

From: Mertz, Prema  
Sent: Monday, July 15, 2002 9:08 AM  
To: STIC-Biotech/ChemLib  
Subject: 09/521,195

Please search SEQ ID NO:2, 4 with DNA databases.

Please search SEQ ID NO:1, 3 with protein databases.

Thanks

Prema Mertz, Ph.D.  
Primary Examiner  
Art Unit 1646  
Crystal Mall 1, Room 10E-01  
United States Patent & Trademark Office  
# (703) 308-4229

Point of Contact:  
Beverly Shears  
Technical Info. Specialist  
CM1 1E05 Tel: 308-4994

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 23:47:19 ; Search time 86.06 Seconds  
(without alignments)  
711.151 Million cell updates/sec

Title: US-09-521-195B-1

Perfect score: 2845

Sequence: 1 MRDYDEVIAFLGEMGFQRL.....KKTROSMETENPKVLITAF 551

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_032802:\*

1: /SID5/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SID5/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SID5/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SID5/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SID5/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SID5/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SID5/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SID5/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SID5/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SID5/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SID5/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SID5/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SID5/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SID5/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SID5/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SID5/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SID5/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SID5/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SID5/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SID5/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SID5/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SID5/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2845	100.0	551	20	AAV01649	A protein with cat
2	2470	86.8	553	20	AAV01651	A protein with cat
3	2470	86.8	553	21	AA20579	Mouse OCTN1 amino
4	2218	78.0	557	20	AAV01650	A protein with cat
5	2218	78.0	557	21	AAV83929	Human carnitine tr
6	2218	78.0	557	22	ABG03029	Novel human diagno
7	2164	76.1	557	20	AAV01652	A protein with cat
8	2164	76.1	557	21	AA20580	Mouse OCTN2 amino
9	2164	76.1	557	21	AAV83930	Mouse carnitine tr
10	2034.5	71.5	564	21	AA20578	Mouse OCTN3 protei
11	857	30.1	548	22	AB26279	Drosophila melanog

12	819	28.8	567	22	AB26354	Drosophila melanog
13	797.5	28.0	577	22	AA278367	Human protein SEQ
14	779.5	27.4	560	21	AA243038	Human ORFX ORF2802
15	779	27.4	584	22	AA200930	Human bone marrow
16	773.5	27.2	561	22	AB260830	Drosophila melanog
17	748.5	26.3	585	22	AA279551	Human protein SEQ
18	703	24.7	555	22	AA279251	Human protein SEQ
19	698	24.5	540	22	AA249401	Murine organic ani
20	695	24.4	574	22	AB258206	Drosophila melanog
21	685.5	24.1	564	19	AA264538	Human liver cell c
22	682	24.0	535	21	AA251249	Rat liver anion tr
23	681.5	24.0	556	17	AA277676	Rat OCT-1 protein.
24	672	23.6	557	22	AA271470	Drosophila melanog
25	671	23.6	565	22	AB211289	Novel human diagno
26	671	23.6	565	22	AB211291	Novel human diagno
27	664.5	23.4	546	22	AA247272	hOAT2A. Homo sapi
28	663.5	23.3	538	22	AA247273	hOAT2B. Homo sapi
29	653.5	23.0	548	21	AA208823	A human organic an
30	653.5	23.0	552	22	AB212030	Human liver-specif
31	651	22.9	542	21	AA292902	Human cerebral org
32	650	22.8	551	21	AA208824	A human organic an
33	647	22.7	542	22	AA247274	hOAT3. Homo sapi
34	637	22.4	536	21	AA292903	Rat cerebral organ
35	636	22.4	561	18	AA244196	Human osteoclast t
36	634.5	22.3	553	22	AA210382	Human osteoclast t
37	630	22.1	537	18	AA244195	Mouse osteoclast t
38	627	22.0	538	22	AB260174	Drosophila melanog
39	624	21.9	607	21	AA212131	Hydrophobic domain
40	621	21.8	578	22	AA206571	Human protein havi
41	614.5	21.6	483	22	AA200982	Human bone marrow
42	606.5	21.3	545	22	AA263553	Mouse organic anio
43	597.5	21.0	551	20	AA268458	Rat organic anion
44	592	20.8	762	22	AB263300	Drosophila melanog
45	586	20.6	542	22	AB263290	Drosophila melanog

#### ALIGNMENTS

RESULT 1	
AAV01649	
ID	AAV01649 standard; Protein: 551 AA.
AC	AAV01649;
XX	
DE	23-JUN-1999 (first entry)
XX	
XX	A protein with cation transporting activity.
KW	Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
KW	heart disease; cancer; anti-tumour drug; anticancer drug.
XX	
OS	Homo sapiens.
XX	
PN	W09913072-A1.
XX	
PD	18-MAR-1999.
XX	
PF	07-SEP-1998; 98WO-JP04009.
XX	
PR	20-MAY-1998; 98JP-0156660.
XX	
PR	08-SEP-1997; 97JP-0260972.
XX	
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX	
PI	Nezu J, Oku A;
XX	
DR	WPI, 1999-215062/18.
XX	
DR	N-PSDB; AA26879.
XX	
PT	Genes homologous with organic cation transporters OCT1 and OCT2,
XX	useful in design of new drugs for treatment of diseases due to
PT	abnormality of the transporter functions

XX Claim 1: Page 41-45; 97pp; Japanese.

XX The present sequence represents a protein with cation transporting  
 CC activity. The genes are significantly homologous with organic cation  
 CC transporters OCT1 and OCT2. The genes may used in drug development,  
 CC particularly in the treatment of diseases due to abnormality of the  
 CC organic cation transporter functions e.g. fatty liver, heart diseases  
 CC and cancers, by controlling such as by inhibition or activation.  
 CC Administration of anti-tumour and anticancer drugs in combination with  
 CC a transporter protein inhibiting agent allows the agents to penetrate  
 CC into the diseased cells to enhance the drug action.

XX Sequence 551 AA;

Query Match 100.0%; Score 2845; DB 20; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 3e-281;  
 Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDYDVIAFLGEMGPORLIFELLSASITPENGNGMSVFLAGTPEHRCRVDAANLSS 60  
 DB 1 mrdydeviaflgempgrliffllsasilpngfngmsvflagtpchrcrvpdaanlss 60  
 QY 61 ABRNNSVPLRLRDGREVPHSCSRYLATIANFSLGLEPGRDVLGOLDESCLDGMEFS 120  
 DB 61 abrnnsvplrlrdgrephscsrylatianfalsalgldpgrdvlgldeescldgmefts 120  
 QY 121 QDVYSTVVTENMLVCEDNMKVPLTTSLEFVGVLLGSFVSGQLSDRGRKNVLFATMAVQ 180  
 DB 121 qdvystvvtenmlvcednmkvpittslefvgvllgsfvsqglsdrgrknvlfatmavq 180  
 QY 181 TGFSEFLQISISWEMFVLFVYVGMGOISNVVAFLIGTEILGKSVRIIFSTLIGCTFPA 240  
 DB 181 tgfseflqisismefvlfvyvgmgoisnvvafligteilgksvriifstlignctfpa 240  
 QY 241 VGYMLPLFAFTRDMRMLLALTVPGVLCVPLMWIPESPRLVISQRRFAEDITOKA 300  
 DB 241 vgymlplfaftrdmrmlalaltvpgvlcvpmlwipesprrlvlsqrrfaeditoka 300  
 QY 301 AKMNTAFAVAFVDSVEELNPLKOOKAFIIDLFTTRNIAIMTMSLLMLLTVSGFALS 360  
 DB 301 akmntafavafvdsveelnplkookafiidlfttrniamtmsllmlltvsqfals 360  
 QY 361 LDAPNLHGDAYLNCFLSALIEIPAYTTAMLLRLTPRRYIAAFLFWGGVLLFTQLVPV 420  
 DB 361 ldapnlhgdaylncflsalieipayttamllrltprryiaaflfwggvllftqlvpv 420  
 QY 421 DYFLSTIGLVMGKFGTSAFMSLVYFTAEIYPTLVNNAVGVTSFASVGSITAPFY 480  
 DB 421 dyflstiglvmgkfgtasmlyvftaelypclvnnavgvtsfsvgsitapfy 480  
 QY 481 LGAYRMLPYIVMGSFLVIGIFLTPFPESLGMTPELTLOMOKVWFRSGKRTDSMET 540  
 DB 481 lgayrmlpyivmgsflvlgifltpfpeslgmtpeletlomokvwfrsgkrtdsmet 540  
 QY 541 EENPKVLTAF 551  
 DB 541 eenpkvltaf 551

# RESULT 2

AAV01651 standard; Protein; 553 AA.

AAV01651;

23-JUN-1999 (first entry)

A protein with cation transporting activity.

Organic cation transporter; OCT1; OCT2; drug development; fatty liver;  
 heart disease; cancer; anti-tumour drug; anticancer drug.

XX Mus. musculus.

XX WO9913072-A1.

XX 18-MAR-1999.

XX 07-SEP-1998; 98WO-JP04009.

XX 20-MAY-1998; 98JP-0156660.

XX 08-SEP-1997; 97JP-0260972.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Nezu J, Oku A;

XX WPI: 1999-215062/18.

XX N-PSDB; AAX26898.

XX Genes homologous with organic cation transporters OCT1 and OCT2,  
 PT useful in design of new drugs for treatment of diseases due to  
 XX abnormality of the transporter functions

XX Claim 1: Page 63-67; 97pp; Japanese.

XX The present sequence represents a protein with cation transporting  
 CC activity. The genes are significantly homologous with organic cation  
 CC transporters OCT1 and OCT2. The genes may used in drug development,  
 CC particularly in the treatment of diseases due to abnormality of the  
 CC organic cation transporter functions e.g. fatty liver, heart diseases  
 CC and cancers, by controlling such as by inhibition or activation.  
 CC Administration of anti-tumour and anticancer drugs in combination with  
 CC a transporter protein inhibiting agent allows the agents to penetrate  
 CC into the diseased cells to enhance the drug action.

XX Sequence 553 AA;

Query Match 86.8%; Score 2470; DB 20; Length 553;  
 Best Local Similarity 84.6%; Pred. No. 5.4e-243;  
 Matches 468; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

QY 1 MRDYDVIAFLGEMGPORLIFELLSASITPENGNGMSVFLAGTPEHRCRVDAANLSS 60  
 DB 1 mrdydeviaflgempgrliffllsasilpngfngmsvflagtpchrcrvpdaanlss 60  
 QY 61 ABRNNSVPLRLRDGREVPHSCSRYLATIANFSLGLEPGRDVLGOLDESCLDGMEFS 120  
 DB 61 abrnnsvplrlrdgrephscsrylatianfalsalgldpgrdvlgldeescldgmefts 120  
 QY 121 QDVYSTVVTENMLVCEDNMKVPLTTSLEFVGVLLGSFVSGQLSDRGRKNVLFATMAVQ 180  
 DB 121 qdvystvvtenmlvcednmkvpittslefvgvllgsfvsqglsdrgrknvlfatmavq 180  
 QY 181 TGFSEFLQISISWEMFVLFVYVGMGOISNVVAFLIGTEILGKSVRIIFSTLIGCTFPA 240  
 DB 181 tgfseflqisismefvlfvyvgmgoisnvvafligteilgksvriifstlignctfpa 240  
 QY 241 VGYMLPLFAFTRDMRMLLALTVPGVLCVPLMWIPESPRLVISQRRFAEDITOKA 300  
 DB 241 vgymlplfaftrdmrmlalaltvpgvlcvpmlwipesprrlvlsqrrfaeditoka 300  
 QY 301 AKMNTAFAVAFVDSVE--ELNPLKOOKAFIIDLFTTRNIAIMTMSLLMLLTVSGYFA 358  
 DB 301 akmntafavafvdsve--elnplkookafiidlfttrniamtmsllmlltvsyfa 360  
 QY 359 LSLDAPNLHGDAYLNCFLSALIEIPAYTTAMLLRLTPRRYIAAFLFWGGVLLFTQLV 418  
 DB 359 lslapnlhgdaylncflsalieipayttamllrltprryiaaflfwggvllftqlv 420  
 QY 419 PVGYFLSTIGLVMGKFGTSAFMSLVYFTAEIYPTLVNNAVGVTSFASVGSITAPFY 478  
 DB 419 pvyflstiglvmgkfgtasmlyvftaelypclvnnavgvtsfsvgsitapfy 480



```
QY 479 YILGAVNRMPLPYIVGSLTVLIGITLFFPESLGMTLPETILEQOKYKWPFSCKKTRDSM 538
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 481 vylgavynrlipyilmgstlvtlglitlffesfgvltlpenlegmqkvrfgcjkkslsv 540

QY 539 ETEENPKVLITAF 551
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 541 dreepkvlitaf 553

RESULT 3
AAB20579
ID AAB20579 standard; Protein: 553 AA.
XX
XX AAB20579;
XX
XX 11-DEC-2000 (first entry)
XX
XX Mouse OCTN1 amino acid sequence.
XX
XX Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
XX identification; regulator; carnitine transport.
XX
XX Mus musculus.
XX
XX WO200046368-A1.
XX
XX 10-AUG-2000.
XX
XX 04-FEB-2000; 2000MO-JP00619.
XX
XX 05-FEB-1999; 99JP-0028406.
XX
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
XX Nezu J, Ose A;
XX
XX WPI; 2000-586982/55.
XX
XX Organic cation transporter gene OCTN3 expressed in testis for
XX identification of regulators of carnitine transport for use as drugs -
XX
XX Example 2; Fig 2; 58pp; Japanese.
XX
XX The present invention describes a mouse organic cation transporter
XX protein (OCTN3). Also described are: (1) a method for screening
XX compounds for their ability to regulate the transport of an organic
XX cation into the cell, by generating a cell expressing OCTN3 at the
XX cell membrane, contacting with the compound and organic cation, and
XX observing the degree of transport of the organic cation; and (2) a
XX method for screening compounds for their ability to be transported into
XX the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
XX membrane, contacting with the compound and observing the degree of
XX transport of the compound. OCTN3 can be used for the identification of
XX regulators of the transport of organic cations (especially carnitine)
XX into cells by OCTN3, for use as drugs. The present sequence represents
XX the mouse OCTN1 amino acid sequence, which is used in an example from
XX the present invention.
XX
XX Sequence 553 AA;
SQ
```

Query Match 86.8%; Score 2470; DB 21; Length 553;  
Best Local Similarity 84.6%; Pred. No. 5, 4e-243;  
Matches 468; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

```
QY 1 MRDYDEVIAFLGEMGPORLIFFLLSASIIIPNGFNMGSVVFLAGTPBHRCPDPAANLSS 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 1 mrdydeviaflgwgpfgrliffllsasllpnqfngmsvvlagtpbhrclvpdvtvlss 60

QY 61 AWRNNSVPLRLRDGRVPHSGRSRYRLATIANESALGLEPGRDVDVLOGLQESCLDGEFS 120
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 61 swrnshpletkdgrvwpqscrrylatiansamslepgqdvdleqldegscldgweyd 120
```

```
QY 121 QDYVLTSTVTEENMLVGEDNMKVPLETTSLFPEVGLIGSFVSGOLSDRFGRKNVLFATMAVO 180
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 121 kdlfstlvtewmlvceddkrpltslftvgvlgstfvsqglsdftgkklvfatmaq 180

QY 181 TGFSEFLQIFSISWMEFTVLVIVMGQISNYVVAFTLGTIELGKSVRIIFSTLGVCTFPA 240
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 181 tgfsefvgifstwenemftvlfaivgmqgsinyvvaflgtelisksvrllfstlgtctffa 240

QY 241 VGYMLPLFAVFIIDRMMLLALTVPQVLCVPLMWTIPESPRMLISQRFREEDITQKA 300
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 241 lgywvplfayfildwmlllaaltlp9lfcvplmwflpesprwljsqrftaeegqllgka 300

QY 301 AKMNNVPAVPIPSVSE--ELNPLKOOKAFITLDEFRTNTAINTISLMLTMLTSVGYFA 358
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 301 akmsnivapagifdplgeinslkqkvllldftrlnactltvmavmwmltsvgyfa 360

QY 359 LSLDAPNLHGDVAYLNCPLSALIEIPAYITAMLLRTPRYITAAVLEFWGCVLFIOLY 418
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 361 lslvplnhgdvynlncfslgllvpayftawlllrtprryllagvlfwggvllllqyv 420

QY 419 PVDYFELSTGLVMGKRGITSAFSMLVPTAAELXPTLYRMMAVGYNSTASRVGSIAPYF 478
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 421 pedyfivsiglvmkgitglsatsmlyvftaelypclivrmavglismastrvslapyf 480

QY 479 YILGAVNRMPLPYIVGSLTVLIGITLFFPESLGMTLPETILEQOKYKWPFSCKKTRDSM 538
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 481 vylgavynrlipyilmgstlvtlglitlffesfgvltlpenlegmqkvrfgcjkkslsv 540

QY 539 ETEENPKVLITAF 551
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 541 dreepkvlitaf 553

RESULT 4
AAY01650
ID AAY01650 standard; Protein: 557 AA.
XX
XX AAY01650;
XX
XX 23-JUN-1999 (first entry)
XX
XX A protein with cation transporting activity.
XX
XX Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
XX heart disease; cancer; anti-tumour drug; anticancer drug.
XX
XX Homo sapiens.
XX
XX WO9913072-A1.
XX
XX 18-MAR-1999.
XX
XX 07-SEP-1998; 98WO-JP04009.
XX
XX 20-MAY-1998; 98JP-0156660.
XX
XX 08-SEP-1997; 97JP-0260972.
XX
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
XX Nezu J, Oku A;
XX
XX WPI; 1999-215062/18.
XX
XX N-PSDB; AAX26880.
XX
XX Genes homologous with organic cation transporters OCT1 and OCT2,
XX useful in design of new drugs for treatment of diseases due to
XX abnormality of the transporter functions
XX
XX Claim 1; Page 51-55; 97pp; Japanese.
XX
XX The present sequence represents a protein with cation transporting
XX activity. The genes are significantly homologous with organic cation
```

CC transporters OCT1 and OCT2. The genes may be used in drug development,  
CC particularly in the treatment of diseases due to abnormality of the  
CC organic cation transporter functions e.g. fatty liver, heart diseases  
CC and cancers, by controlling such as by inhibition or activation.  
CC Administration of anti-tumour and anticancer drugs in combination with  
CC a transporter protein inhibiting agent allows the agents to penetrate  
CC into the diseased cells to enhance the drug action.

XX Sequence 557 AA;

Query Match 78.0% Score 2218; DB 20; Length 557;  
Best Local Similarity 75.9%; Pred. No. 2,8e-217;  
Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

QY 1 MRDYDEVIAFLGEMGPORLIFLLSASIIIPNGFNGSVFLAGTPHRCRPDAANLS 60  
DB 1 mrdydeviaflgempgrliffllsasiiipngfngsvflagtpghrcrpdanls 60  
QY 61 AMRNSVPLRLRDGRVPHSCSRRLATIANFSALGLEPRDVLGQLQESCLDGMERS 120  
DB 61 awrnhvplrlrdgrevphscrrlatianfsalgleprdvlgqlqescldgwefs 120  
QY 121 QDYLSTVVTENMLVCEENMKVPLTSLFVGVLSFVSGQLSDRFGKKNVLFATMAVQ 180  
DB 121 qdylstvtvtenmlvcednmkvpלטלסלפגולסדפגקקנולפאטמאוק 180  
QY 181 TGFSEFLQISISWEMETVLEFVIVGMQISNYVAFILGTEILGKSVRIESTLGVCTPFA 240  
DB 181 tgfseflqisisknfemfvlvlyvgmgisnyvaflgteilgksvriestlgyvcifya 240  
QY 241 VGYMLPLFAFIRDMRMLLATLVPGVLCVPLWMPFIPSPRMLISQRRFEADIIOKA 300  
DB 241 fgymlplfayfirdwrmlllatlvpgvlcvalwmpfipspwllsqrrfeaeavllrka 300  
QY 301 AKNNNTAVPAVIFD--SVEELNPLKQOKAFILDFRTNRNIAIMTMSLLMLTSVGIFA 358  
DB 301 akangivvpstlfdpselqdksskqshnldlrltwlmvtsimlmwllsvgyfg 360  
QY 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAMLLRPLPRRYITAAVLEWGGVLLFIOLV 418  
DB 359 lsldapnlhgdaylncflsalieipayitamllrplprryitaa vlewgvgvllfiolv 418  
QY 419 PVDYFRLSTGLVLMGKFGITSAFSLVYFAELYPVLRNNAVGVTASRVGSIIAPYF 478  
DB 419 pvdylfslstglvmlkgfirtsafslvlyfaelypvlnnmgvgsstasrvgsiispyf 480  
QY 479 VYLGAVNRMLPYIVMGSLTVLIGITLFFPESLGMTLPETLEQOKKVRFRSGK---KTR 535  
DB 479 vylgavnrmlpyivmgsltvligitlffpesslgmtlpetleqokkvrfrsgk---ktr 535  
QY 536 DSMETEENPKVL-ITAF 551  
DB 541 mlkdqgerptllkstaf 557

RESULT 5  
ID AAY83929 standard; Protein: 557 AA.  
XX AAY83929;  
XX  
XX 05-JUL-2000 (first entry)  
XX  
XX Human carnitine transporter protein OCTN2.  
XX  
XX Organic cation transportation; human; carnitine transporter protein;  
XX OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;  
XX juvenile visceral steatosis.  
XX Homo sapiens.  
XX  
XX WO200014210-A1.  
PN

XX  
PD 16-MAR-2000.  
XX  
XX 07-SEP-1999; 99WO-JP04853.  
PF  
XX 07-SEP-1998; 98JP-0252683.  
PR  
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
PA  
PI Nezu J, Oku A;  
XX  
XX WPI: 2000-256966/22.  
DR N-PSDB: AAA09889.  
XX

PT Systemic carnitine deficiency gene OCTN2 encoding part of organic  
PT cation transporter, useful as diagnostic tool -  
PS Example 1; Page 36-40; 106pp; Japanese.

CC This sequence represents the human carnitine transporter protein OCTN2.  
CC The coding sequence can be used as a target for diagnosis of systemic  
CC carnitine deficiency by detecting the presence of mutations in the  
CC sequence, especially seen in the disease juvenile visceral steatosis  
CC (JVS). The wild type OCTN2 gene can be used in the gene therapy of the  
CC disease state.

XX Sequence 557 AA;

Query Match 78.0% Score 2218; DB 21; Length 557;  
Best Local Similarity 75.9%; Pred. No. 2,8e-217;  
Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

QY 1 MRDYDEVIAFLGEMGPORLIFLLSASIIIPNGFNGSVFLAGTPHRCRPDAANLS 60  
DB 1 mrdydeviaflgempgrliffllsasiiipngfngsvflagtpghrcrpdanls 60  
QY 61 AMRNSVPLRLRDGRVPHSCSRRLATIANFSALGLEPRDVLGQLQESCLDGMERS 120  
DB 61 awrnhvplrlrdgrevphscrrlatianfsalgleprdvlgqlqescldgwefs 120  
QY 121 QDYLSTVVTENMLVCEENMKVPLTSLFVGVLSFVSGQLSDRFGKKNVLFATMAVQ 180  
DB 121 qdylstvtvtenmlvcednmkvpלטלסלפגולסדפגקקנולפאטמאוק 180  
QY 181 TGFSEFLQISISWEMETVLEFVIVGMQISNYVAFILGTEILGKSVRIESTLGVCTPFA 240  
DB 181 tgfseflqisisknfemfvlvlyvgmgisnyvaflgteilgksvriestlgyvcifya 240  
QY 241 VGYMLPLFAFIRDMRMLLATLVPGVLCVPLWMPFIPSPRMLISQRRFEADIIOKA 300  
DB 241 fgymlplfayfirdwrmlllatlvpgvlcvalwmpfipspwllsqrrfeaeavllrka 300  
QY 301 AKNNNTAVPAVIFD--SVEELNPLKQOKAFILDFRTNRNIAIMTMSLLMLTSVGIFA 358  
DB 301 akangivvpstlfdpselqdksskqshnldlrltwlmvtsimlmwllsvgyfg 360  
QY 359 LSLDAPNLHGDAYLNCFLSALIEIPAYITAMLLRPLPRRYITAAVLEWGGVLLFIOLV 418  
DB 359 lsldapnlhgdaylncflsalieipayitamllrplprryitaa vlewgvgvllfiolv 418  
QY 419 PVDYFRLSTGLVLMGKFGITSAFSLVYFAELYPVLRNNAVGVTASRVGSIIAPYF 478  
DB 419 pvdylfslstglvmlkgfirtsafslvlyfaelypvlnnmgvgsstasrvgsiispyf 480  
QY 479 VYLGAVNRMLPYIVMGSLTVLIGITLFFPESLGMTLPETLEQOKKVRFRSGK---KTR 535  
DB 479 vylgavnrmlpyivmgsltvligitlffpesslgmtlpetleqokkvrfrsgk---ktr 535  
QY 536 DSMETEENPKVL-ITAF 551  
DB 541 mlkdqgerptllkstaf 557

RESULT 6  
 ABG03029  
 ID ABG03029 standard; Protein; 557 AA.  
 AC  
 XX ABG03029;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #3020.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS  
 XX Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YF;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS67216.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 33388; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

Query Match	78.0%;	Score 2218;	DB 22;	Length 557;
Best Local Similarity	75.9%;	Pred. No. 2.8e-217;		
Matches 423; Conservative	57;	Mismatches 71;	Indels 6;	Gaps 3

61 AMRNNSVPLRLDRGKREVPHSCRRRLATIANFSAALGEPGRVDVIGQLQEOESCIDGWEFS 120

```

Db      61 awrnhvplrltdgreephscrtxyrtatiansfaqlgplepdrtdlqglegescldyefis
OY      121 QDYYLSTVWTENWLVCEBDMKVPPLTSLFVGVVLGSPVSGOLSDRFGRKAVLFAIMAVO
Db      121 qdyvylstflvtemwlvceoddmkaptlslftfyvllgslfsgqlsdrfgxknhvlftvmqg
OY      181 TGSEFLOISISEMEFTVLFVYVKGGOISNVVAFLIGIEILGKKVRITFSTLVCFFFA
Db      181 tgsftlqfiskfemlvylftvlymgqgslmgyaafvlgelilgkavrilftstlqvclfy
OY      241 VGNMLPLPRAVYFIRBMRHMLLALVPGVLCVPLMWPSPSPMLISQRRFBEADIDQKA
Db      241 fgyvmlplrlaylrlfrowrmllyallcmppylcxaawllfpeprrvllsgrfseaaellrka
OY      301 AKNNNTAHPAVTFFD--SVEELNPLKQOKAFLTDLFRTRNIAIMTMSLLMLNMTSVGYFA
Db      301 akanglvpsrlftfpseqlgskkqgshllqlrltwnlrmvtlmslmlwmtlsvgyf
OY      359 LSLDAPMLHSDAVLNCFLSALIEPRVATVMTMLRLTPRXYTIAAVLFMGGVLFLFIQOLV
Db      361 lsltdpmlhgsdavlncfslsalieprvatvmtmlrltpxytyiaavlfmggvlflfiqolv
OY      419 PVDFYELSLGLVNLGKFGITSAFSMLYVTFDAELPTLVLRNMAVGVTSTASRVGSIITAPYF
Db      421 pddlylatvlnwvkgfytaasfsmvytaelrpylvtvrimgvyvsslastrlgsllspyf
OY      479 VYLGAVRMLPPIYVMSGLVLVLGIRTFEPFSPSGMTLPETLEBOKVYVWFSKGF---KTR
Db      481 vylygavrfpilylmgslstlftlaaltlftlfpesfgtrpdlldlqdmrlrvymkhrtgshtr
OY      536 DSMETENPKVL-TIAF 551
Db      541 mlkgdgpercltkstaf 557

```

RESULT 7  
AA01652  
ID AA01652 standard; Protein: 557 AA.  
XX  
XX AA01652;  
AC  
XX  
DT 23-JUN-1999 (first entry)  
XX  
XX  
DE A protein with cation transporting activity.  
XX  
KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver  
KW heart disease; cancer; anti-tumour drug; anticancer drug.  
XX  
XX OS Mus musculus.  
XX  
XX PN W09913072-A1.  
XX  
PD 18-MAR-1999.  
XX  
PE 07-SEP-1998; 98WO-JP04009.  
XX  
XX 20-MAY-1998; 98JP-0156660.  
PR 08-SEP-1997; 97JP-0260972.  
XX  
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX  
PI Nezu J, Oka A;  
XX  
XX MPI: 1999-215062/18.  
DR N-PSDB; AAX26902.  
XX  
PT Genes homologous with organic cation transporters OCT1 and OCT2,  
PT useful in design of new drugs for treatment of diseases due to  
PT abnormality of the transporter functions  
XX  
PS Claim 1; Page 75-79; 97JP: Japanese.





CC Into cells by OCTN3, for use as drugs. The present sequence represents  
CC mouse OCTN3.  
XX  
SQ Sequence 564 AA;

50 Sequence 564 AA;

Query Match	71.5%;	Score 2034.5;	DB 21;	Length 564;
Best Local Similarity	69.0%;	Pred. No. 1.5e-198;		
Matches 380;	Conservative 75;	Mismatches 85;	Indels 11;	Gaps 3;

QY	1	MROVDIEIAFEGE	PPORLRIIEFLFASITIPNCFNM	SVFLAGTPEHRCRPA	DAA	NTSS
Db	1	mlqydetalfi	geqgtcgrliffllsa	slpnngfcglsavf	ltaipenrcr	tpdvnlts
QY	61	AMRNNSVPLRL	DGREGVPHSCSR	RYRLATTANFSA	LGEPDVGDLG	EQESCUDGWEF
Db	61	awrnhsipmet	kdqpevpqkrcrry	latlanfseagl	ieprgrvdl	leqlegencldgwey
QY	121	ODVLSIVTV	EMNLVCEDDNKK	PLRTNLSLFPVGLGS	FPVSGDLSR	PGKNVLFATMAV
Db	121	kdllfistvie	ewdlvvcddwk	kapltsctfityg	vllygsfissglsdr	fgkmlfltmam
QY	181	TGFSFLDIFS	ISWEMETVLE	EVIMGQISNYVA	FLIGLTELIGK	SVRIIFS
Db	181	tgfsfigv	favnfemfclly	tylvgmghisny	aafvlgtemisk	svrllfclgyclfeff
QY	241	VGVMLDLPL	EAFYTRDRM	KMLLATATVGC	VLCPVLMWIPES	PRMLISQRRRAEDII
Db	241	fgtmvlp	lflfayflfiew	rtrlllatlrlp	vgvlegalmwflipes	prwllsqgkikeaevlirk
QY	301	AKMNNTV	AVPPIEVSE	VELNPL-----	KQOKAFILDE	PRFNINIMTMSLLM
Db	301	aklngv	apstlfdq	-sencklq	dsckkpgqshjy	ldvtrtpnrlfltlmsllwlts
QY	355	GYPFLSIDA	ENLHGDA	LKNCFLSAL	LEIPAYITAM	LLFLTERPRTITAAVLF
Db	360	gyfglsld	pcnlnghnyc	fllaavep	ayvlawllqhv	srlysmagsllgsvll
QY	415	IQLPVV	YELFESTL	VWLKFGF	GTSAF	SMUYETAE
Db	420	vqlpsal	hlylsc	tlwmgkfg	ltsaysmwy	ylaelyplvtrmngvys
QY	475	APRYVY	IGATNRM	PIYVMS	GLTVLIGIT	FLPSPESLG
Db	480	spyfvy	igaydr	trpyllms	lfltalitl	flfressyvsldetidemg
QY	535	RDSMETEN	PK	545		
Db	536	rgslsk	kspsk	546		
RESULT	11					
ID	ABB62279					
XX	ABB62279					
AC	ABB62279;					
DT	26-MAR-2002					
XX	(first entry)					
DE	Drosophila melanogaster					
XX	polypeptide seq					
OS	ID NO 13629.					
XX						
KW	Drosophila: developmental biology; cell signalling; insecticide;					
XX	pharmaceutical.					
OS	Drosophila melanogaster.					
PN	WC000171042-A2.					
XX						
PD	27-SEP-2001.					
XX						
FE	23-MAR-2001; 2001WC-US09321.					
XX						
PR	23-MAR-2000; 2000US-191637P.					

PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PMD, Myers EW;  
PI WPI; 2001-656860/75.  
DR N-PSDB; ABL06382.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 13629; 21pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57137-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcc\_sequences.  
XX  
XX Sequence 548 AA;

Query Match	30.1%;	Score 857;	DB 22;	Length 548;
Best Local Similarity	35.7%;	Pred. No. 1.9e-78;		
Matches 201; Conservative	98;	Mismatches 208;	Indels 56;	Gaps 10

```

OY 4 YDEYIAFLGEMBPORLIEFLLSASITPNNGENGSVYFLAGTREPBRICVP---DAANLS 59
Db 3 yddvithllegfgrgykxrluyllclclpaivcshlagfyllakpofccalpryngslyels 62
OY 60 SAMRNSVPLRLRDGNEVHSSCRRLATIANFSLGLEPCROYDL-----GQL----- 10
Db 63 phlmnlsypene-----csy-----dvdyleeylngsiprsn 97
OY 109 EOESLDGMEFSDOVYLSIVTVEKMLICEDMMKVPFLTSLFEVGLLGSVSGQSDRPG 16
Db 98 etkc-syyydatskylmsaavtemnlvsrsllsatsdsifmgyvllgslifgmsdklg 15
OY 169 RKNVLEPMNAVQGSFSLQIFSIISMEFTVLVYVNGQISNVVAFTIGETILKSVRI 22
Db 157 rkpfefastvllqllfgvlaavapefsyrlsrmvlatgsvfayvalalemgssyr- 21
OY 229 IESTLGCFFPFGVWLLPLFAYFLRDBRMILLALTPGVYLCVPLMMFELPESRMLISOR 28
Db 216 lfavaamqmfsvygmthtagfayflndwrtqialclpyllficywllpresarvllimg 27
OY 289 RFRRAEDIIOKAKAMNNTAVPAVIDFS-VEEELNLPKO-----KAFILDFTFRINA 33
Db 276 rkgeatfillekaekvneymelveyqlvdeaeakkkqdeaaasqpaetvdllyrplnr 33
OY 340 IMTMSLLMLMTLSVGYFALSLDABNLHGDAYLNCFSALIEIPATYTMALLRLPLRRY 39
Db 336 rkclllffdfvnsygyylswntnnlgsnqlvfmisgaelpyrlillfltnwrgs 39
OY 400 IIAAVLEWGGVLLFLQIVPVPVYFLSLGIVMLCKFGTSTAFSLMYEPAELYPTRLVRM 45
Db 396 llcgtmvaaisllactipysdmnlvacamicklatsygytlfysseqprvtvrv 45
OY 460 AVGTSTARSGVSIIRPEVYVGAANRLPIYVNGSLTVLIGITFLFEFSLMTLEPL 51
Db 456 glgssmvaavvgvllapyrklllgelwtrpplliogalsltagllsllpctlnkmpeti 51
OY 520 EOMCKVMEFRSGAKTRDSMETEE 542
Db 516 edgenf-----qkppapgetae 533

```

RESULT 12  
ID ABB68354 standard; Protein: 567 AA.  
AC ABB68354;  
DE 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 31854.  
XX DE  
XX Drosophila: developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX PR 23-MAR-2000; 2000US-191637P.  
XX PR 11-JUL-2000; 2000US-0614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX DR MPI: 2001-656860/75.  
XX DR N-PSDB: ABL12457.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX PS Disclosure: SEQ ID NO 31854; 21pp + Sequence Listing; English.  
XX CC The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
XX sequences (AB101840-AB116175) and the encoded proteins  
XX (ABB57737-ABB72072).  
XX CC The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 567 AA:  
Query Match 28.8%; Score 819; DB 22; Length 567;  
Best Local Similarity 34.0%; Pred. No. 1.5e-74;  
Matches 194; Conservative 112; Mismatches 198; Indels 66; Gaps 12;  
QY 4 YDEVIAFLGWMGPFQRIFFLLSASITPENGNGMSVFLAGTEPHRCRPV---DANIS 59  
DB 3 ydealllgdgrqkilylilictslpafhklagfllakpdlcafpfengssydlp 62  
QY 60 SAMNNSVPLRLRDGREVPHSCSRVRLATIANFSALGLEPGRVVDL-----GQL----- 108  
DB 63 thlwlylpener-----ceyy-----dvdyleeylngslprsn 97  
QY 109 EQESCGLDGEFSQDVYSTVVTENMLVCEDNMKVPLTTSLFYGVYLGVSQQLSDRFG 168  
DB 98 ektcc-syydydrskynsawtemwlvcygrdmatadsllfmjgvlgsivfgqlsdxyg 156  
QY 169 RKNVLPFPMNVQGESFLQJFSISWEMFTVLF--VIWGMGOISNVVAFILGIEIIGKSV 226  
DB 157 tkpillsivqvlfgvla--gvapeyftlytarlmvgatstgyflvayvvanemvypdk 214

QY 227 RIIFSTLGCTFFAFVGYMLPLFAFYIRDMRMILLATVPGVLCVLMWFIDESPRLIS 286  
DB 215 r-lyaglfvmmffsvgfmlavfayfvdwrvqialtlpgllfvmfywldipesarwll 273  
QY 287 QRRFEAREEDITQKAANKNTAVP----AVTFDSVEELNPLKOOK-----AFILD 331  
DB 274 kyrkdcaianmukaarfinkveisdelselldegenseekakqkledgeldegppsvvd 333  
QY 332 LFRTRNIAIMTISLLIMLTSGVFPALSDADVNLGDVATNCFSLAEIPYIRYAWLL 391  
DB 334 lfcpnlrktlllflldwlvtsgyyglswtsnlggnvllnvisgaveipayflillt 393  
QY 392 LRTLPRRYIIAAVLFWGCVLFIQVVDVYFSLGVLVIGKFGITSASFMLVFTAEU 451  
DB 394 lntwgrtsllcglvmaglllatvllpqrmhcllvacamlgklatlasytylfsaeq 453  
QY 452 YPTLVKMAVGVSTASRVGSIIAPFYVYGATNRMPIYVMSLVYLGIFLFPFESL 511  
DB 454 fptvrnvajgaasmvarlsqmmapiinflatiwkplllcgsltlvqllslilpeth 513  
QY 512 GMTLPETLEQMKVKWFRSGKTRDSMETE 541  
DB 514 nkpmlectadg-----rfgkktkadvyile 538  
RESULT 13  
ID AAM78367 standard; Protein: 577 AA.  
AC AAM78367;  
DE 06-NOV-2001 (first entry)  
XX Human protein SEQ ID NO 1029.  
XX DE  
XX KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorder; arthritis; inflammation.  
XX OS Homo sapiens.  
XX PN WO200157190-A2.  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US04098.  
XX PR 03-FEB-2000; 2000US-0496914.  
XX PR 27-APR-2000; 2000US-0560875.  
XX PR 20-JUN-2000; 2000US-0598075.  
XX PR 19-JUL-2000; 2000US-0620325.  
XX PR 01-SEP-2000; 2000US-0654936.  
XX PR 15-SEP-2000; 2000US-0663561.  
XX PR 20-OCT-2000; 2000US-0693325.  
XX PR 30-NOV-2000; 2000US-0728422.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Liu C, Dymnac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
XX Xue AJ, Yang Y, Wehrman T, Goodrich R;  
XX WPI: 2001-476283/51.  
XX N-PSDB: AAK51500.  
XX PT Nucleic acids encoding polypeptides with cytokine-like activities,  
XX useful in diagnosis and gene therapy -  
XX Claim 20; Page 3266-3267; 6221pp; English.  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation of which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haemopoietic regulatory  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

XX Sequence 577 AA:

Query Match 28.0%; Score 797.5; DB 22; Length 577;

Best Local Similarity 32.5%; Pred. No. 2,4e-72; Mismatches 221; Indels 49; Gaps 9;

Matches 187; Conservative 118; Mismatches 221; Indels 49; Gaps 9;

QY 2 RDYEVIAFLGEMPPORLIEFLLSASTIPNGFNGMSVFLAGTPEHRCRP----- 53  
DB 4 Rlfegldhvhgfrfgrfyficalfniscgihylasvfmgtvphvcprpnsqvfvf 63  
QY 54 -----DAANLSAMRNNSVPLRLRDGREVPHS--CSRRLATIANFSLGPEGRDV 103  
DB 64 hnhmswledtgallssgqkdyvtvqlqgealwelsrstrkr---entslgye----- 115  
QY 104 DLGQLEQESCLDGWESODVYLSTVVTWMLVCEDNMKVPLTSLFVGVLLGSFVSGQL 163  
DB 116 ytskckefpcvdgyldqgtwktavtgmnlvcdrkwlamlqplfmgyvllsgvtfygf 175  
QY 164 SDPFGKKNVLFPMVAVQGESFLQIFESISWEMFTVLFIVMGQISNVVAFILGTEL 223  
DB 176 sdllgrtvtvawssmflfgiaafavdytlfmaarffimvasgylvagfvyvmefig 235  
QY 224 KSVRIESTLGVCTFEFVAGYMLPLFAYFIRDW--RMLLALTPGVLCVPLMFIPE 280  
DB 236 mksr-lwasvnhlshffavgtllvaltylvrtwvlygmhlstvcvflcc---wvlpet 291  
QY 281 PRLILSORPREAEDIIQAAKMNMTA-----VPAVFEDEVBELPLKOOKAFILDLR 334  
DB 292 pvlasegryeeagkivdimaakasscklssllsldlqgvpnspevqkhnlsyly 351  
QY 335 TRNIAIMTMSLLMLWLTSGVGFALSLDAPMLHGDVYNCLSLAJETPAYTAMLLRT 394  
DB 352 nweilkrcltvtwllwftgslgfyfslnsnvgneynlflgyveipaytfcvclamd 411  
QY 395 LPRRYIIAIVLFMGGGVLLFLQLVVDVYFSLIGLVMLGKRGISAFSMLVFTAEIYPT 454  
DB 412 vgrtvtvaystlcsalagvnmvlpqknyllgvtamagktaigafiglylaaelypc 471  
QY 455 LVANNAVGTSTASRVGSIADPYVYLGAYNMLPYIVMGSLFVYLIGITLFFPESLG 514  
DB 472 lvsrlavgsmsvcrslasllapfsvdlsiswffipqlftvgmalisgyltlklpeltg 531  
QY 515 LPELTLEOMKVKMFRSGKTRDSMTEDEPKYLIT 549  
DB 532 laltweeakl-----esenesksklllt 556  
RESULT 14  
ID AAB43038  
ID AAB43038 standard: Protein; 560 AA.  
XX AAB43038;  
XX AC  
XX AAB43038;  
XX 08-FEB-2001 (first entry)  
XX DE Human ORFX ORF2802 polypeptide sequence SEQ ID NO:5604.  
XX Human; open reading frame; ORFX; detection: cytostatic; hepatotropic;

KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritis; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antihypoid;  
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.

XX Homo sapiens.  
XX OS  
XX WO200058473-A2.  
XX PN  
XX 05-OCT-2000.  
XX PD  
XX 31-MAR-2000; 2000WO-US08621.  
XX PR  
XX 31-MAR-1999; 99US-0127607.  
XX PR 02-APR-1999; 99US-0127636.  
XX PR 05-APR-1999; 99US-0127728.  
XX PR 30-MAR-2000; 2000US-0540763.  
XX PA  
XX (CURA-) CURAGEN CORP.  
XX PI  
XX Shinkets RA, Leach M;  
XX WPI; 2000-602362/57.  
XX DR N-PSDB; AAC77247.  
XX XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
XX useful for treating e.g. cancers, proliferative disorders,  
XX neurodegenerative disorders and cardiovascular disease -  
XX Claim 11; Page 4791-4792; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritis; coagulant; vasotropic;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antihypoid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 560 AA:

Query Match 27.4%; Score 779.5; DB 21; Length 560;

Best Local Similarity 32.9%; Pred. No. 1.6e-70; Mismatches 184; Conservative 113; Mismatches 214; Indels 49; Gaps 9;

QY 17 FORLIFLLSASTIPNGFNGMSVFLAGTPEHRCRP-----DAANLS 59  
DB 2 fgrvlyficalfniscgihylasvfmgtvphvcprpnsqvfvfnhmswledtgall 61





Wed Jul 17 08:31:01 2002

us-09-521-195b-1.rag

Search completed: July 17, 2002, 02:03:06  
Job time: 8147 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 17, 2002, 00:27:44 ; Search time 37.9 Seconds

(without alignments)  
355.106 Million cell updates/sec

Title: US-09-521-195b-1

Perfect score: 2845  
Sequence: 1 MDYDEVIAFLGEMGFQRL.....KTRDSMETENPKVLITAF 551Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCtus.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	702	24.7	555 3	US-08-501-572-3	Sequence 3, Appli
2	702	24.7	555 3	US-09-040-444-3	Sequence 3, Appli
3	687.5	24.2	556 3	US-08-501-572-1	Sequence 1, Appli
4	687.5	24.2	556 3	US-09-040-444-1	Sequence 1, Appli
5	680	23.9	553 3	US-08-501-572-2	Sequence 2, Appli
6	680	23.9	553 3	US-09-040-444-2	Sequence 2, Appli
7	630	22.1	537 2	US-08-647-397-2	Sequence 2, Appli
8	267	9.4	520 4	US-08-964-127-2	Sequence 2, Appli
9	267	9.4	520 4	US-09-496-692-2	Sequence 2, Appli
10	234.5	8.2	494 4	US-09-031-392-5	Sequence 5, Appli
11	234.5	8.2	494 4	US-09-299-549-5	Sequence 5, Appli
12	234.5	8.2	494 4	US-09-610-417-5	Sequence 5, Appli
13	222	7.8	492 2	US-08-355-844-3	Sequence 3, Appli
14	222	7.8	492 2	PCT-US95-16126-3	Sequence 3, Appli
15	216	7.6	493 5	US-09-031-392-10	Sequence 10, Appli
16	216	7.6	493 4	US-09-299-549-10	Sequence 10, Appli
17	216	7.6	493 4	US-09-610-417-10	Sequence 10, Appli
18	213.5	7.5	524 2	US-08-928-692-12	Sequence 12, Appli
19	203	7.1	286 4	US-08-964-127-4	Sequence 4, Appli
20	203	7.1	286 4	US-09-496-692-4	Sequence 4, Appli
21	196	6.9	500 4	US-09-031-392-7	Sequence 7, Appli
22	196	6.9	500 4	US-09-299-549-7	Sequence 7, Appli
23	196	6.9	500 4	US-09-610-417-7	Sequence 7, Appli
24	181	6.4	509 2	US-09-031-392-6	Sequence 6, Appli
25	181	6.4	509 4	US-09-299-549-6	Sequence 6, Appli
26	181	6.4	509 4	US-09-610-417-6	Sequence 6, Appli
27	174.5	6.1	584 2	US-08-928-692-13	Sequence 13, Appli

28	172	6.0	383 2	US-09-031-392-3	Sequence 3, Appli
29	172	6.0	383 4	US-09-299-549-3	Sequence 3, Appli
30	172	6.0	383 4	US-09-610-417-3	Sequence 3, Appli
31	169.5	6.0	488 2	US-08-928-692-11	Sequence 11, Appli
32	168.5	5.9	488 2	US-08-928-692-10	Sequence 10, Appli
33	162	5.7	563 2	US-09-031-392-2	Sequence 2, Appli
34	162	5.7	563 4	US-09-299-549-2	Sequence 2, Appli
35	162	5.7	563 4	US-09-610-417-2	Sequence 2, Appli
36	158.5	5.6	109 2	US-08-647-397-4	Sequence 4, Appli
37	156.5	5.5	534 2	US-09-031-392-4	Sequence 4, Appli
38	156.5	5.5	534 4	US-09-299-549-4	Sequence 4, Appli
39	156.5	5.5	534 4	US-09-610-417-4	Sequence 4, Appli
40	149.5	5.3	322 4	US-08-964-127-6	Sequence 6, Appli
41	149.5	5.3	322 4	US-09-496-692-6	Sequence 6, Appli
42	110	3.9	473 1	US-08-597-236-13	Sequence 13, Appli
43	110	3.9	473 1	US-08-746-682A-13	Sequence 13, Appli
44	110	3.9	548 3	US-08-903-139B-9	Sequence 9, Appli
45	109.5	3.8	834 2	US-08-677-734A-9	Sequence 9, Appli

## ALIGNMENTS

RESULT 1  
US-08-501-572-3  
; Sequence 3, Application US/08501572  
; Patent No. 6063623  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; TITLE OF INVENTION: Transport protein which effects the  
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/Or Pharmaceuticals,  
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan,Henderson,Farbow,Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; City: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/501,572  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toohay, Kimberlin M  
; REGISTRATION NUMBER: 35,391  
; REFERENCE/DOCKET NUMBER: 02481.1453-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 555 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-501-572-3

Query Match 24.7%; Score 702; DB 3; Length 555;  
Best Local Similarity 33.7%; Pred. No. 5.6e-65;  
Matches 187; Conservative 93; Mismatches 229; Indels 46; Gaps 14;  
OY 5 DEVIAFLGEMGFQRLIF---LLSASTIIFNGMGSVFLAGTPEHRCRVPDAANIS-- 59

```

Db 6 DDLHEGGEHFHFKOMFLLALLSATFAP---IYGVIFLGFTPDHCRSPGVAELSLR 62
QY 60 SAMR-----NSVPLRLRDGREVPHSCSRYLRA-TIANFSALGLEPGADVLOGLEQESC 113
Db 63 CGWSPAELNLYTPGPGPAGASPPROCRRYEVDMMNSTFDCVDPLASLDITNRSRLPLGPC 122
QY 114 LDGMEFSODVYLTSTVTENMLVCEDNMKVPPLTSLFPGVLLGSFVSQGLSDRGKRNVL 173
Db 123 RCGWY--ETPGSSIVTEFNLYCANSMWLDLFQSSVNVGFFIGSKSIGIADREGKRLCL 180
QY 174 FATMAVQTGFSEFLQIFISMEFTVLFIYVGMGOISNYVAIFLITELIGKSVRIIFSTL 233
Db 181 LTTVLINAAAGVYLAISPTIYMWLIFRLIOGLVSKAGWLIIGYILITEFGGRYR---RTV 237
QY 234 GVC--TFEAVGYMLLPLFAFYFIROMRMLLALYVGVLCVPLMWFIPESPRLISQRRFR 291
Db 238 GIFYQVAVTVGLVLGAVYALPHWRMLQFTVALPNEFFLLYWCIPESPRLISQNKNA 297
QY 292 EAEDIIQAKAKMNTAVPAVI-----FDSVEELNPLKOKAFILDLFRTNIAIMTMS 345
Db 298 EAMRIKRIAKKNGKSLPASIORLREBETGKLN-----SFLDLYRTPIQIKHTMIL 351
QY 346 LLLMLITSVGYFALSIDAPNLHGD-AYLNCFLSALIEIPAYITAMLLRLTPRYITAAV 404
Db 352 MYNFTSSVLYOGLIMH-GLAGDNIFYLDFEYSALVEPPAFMILLIDRIGRRYPMAAS 410
QY 405 LFMGGVLLFIQLVDPDYVFLSIGVLMGKRGITSAFSMLYFAELYPTLRMNAVVT 464
Db 411 NMVAGACLASVFIPTDQWLKIIISCLGRMGITMAEIVCLVNAELPTFIRNLGVHIC 470
QY 465 STASRVGSIAPYFYLYGAVNRM-----LPIYVGSILTVLIGITLFPFESLGTLPET 518
Db 471 SSMCDIGIITPFLVY-----RLTNIMLELPLMVGVLGIVAGGLALLPETKKAALPET 525
QY 519 LQOMOKVWRBSGRK 533
Db 526 IEEAENMORPRKKE 540

RESULT 2
US-09-040-44-3
; Sequence 3, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; City: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-040-44-3

Query Match 24.7%; Score 702; DB 3; Length 555;
Best Local Similarity 33.7%; Pred. No. 5.6e-65;
Matches 187; Conservative 93; Mismatches 229; Indels 46; Gaps 14;

QY 5 DEVIAFLGEMWGPQRLIFF---LISAIIPNGFNGSVYFLAGTPPHKRCRVDAANLS-- 59
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 DDLHEGGEHFHFKOMFLLALLSATFAP---IYGVIFLGFTPDHCRSPGVAELSLR 62
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 SAMR-----NSVPLRLRDGREVPHSCSRYLRA-TIANFSALGLEPGADVLOGLEQESC 113
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 CGWSPAELNLYTPGPGPAGASPPROCRRYEVDMMNSTFDCVDPLASLDITNRSRLPLGPC 122
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 LDGMEFSODVYLTSTVTENMLVCEDNMKVPPLTSLFPGVLLGSFVSQGLSDRGKRNVL 173
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 RCGWY--ETPGSSIVTEFNLYCANSMWLDLFQSSVNVGFFIGSKSIGIADREGKRLCL 180
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 FATMAVQTGFSEFLQIFISMEFTVLFIYVGMGOISNYVAIFLITELIGKSVRIIFSTL 233
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 LTTVLINAAAGVYLAISPTIYMWLIFRLIOGLVSKAGWLIIGYILITEFGGRYR---RTV 237
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 GVC--TFEAVGYMLLPLFAFYFIROMRMLLALYVGVLCVPLMWFIPESPRLISQRRFR 291
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 GIFYQVAVTVGLVLGAVYALPHWRMLQFTVALPNEFFLLYWCIPESPRLISQNKNA 297
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 292 EAEDIIQAKAKMNTAVPAVI-----FDSVEELNPLKOKAFILDLFRTNIAIMTMS 345
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 EAMRIKRIAKKNGKSLPASIORLREBETGKLN-----SFLDLYRTPIQIKHTMIL 351
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 346 LLLMLITSVGYFALSIDAPNLHGD-AYLNCFLSALIEIPAYITAMLLRLTPRYITAAV 404
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 MYNFTSSVLYOGLIMH-GLAGDNIFYLDFEYSALVEPPAFMILLIDRIGRRYPMAAS 410
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 405 LFMGGVLLFIQLVDPDYVFLSIGVLMGKRGITSAFSMLYFAELYPTLRMNAVVT 464
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 NMVAGACLASVFIPTDQWLKIIISCLGRMGITMAEIVCLVNAELPTFIRNLGVHIC 470
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 465 STASRVGSIAPYFYLYGAVNRM-----LPIYVGSILTVLIGITLFPFESLGTLPET 518
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 471 SSMCDIGIITPFLVY-----RLTNIMLELPLMVGVLGIVAGGLALLPETKKAALPET 525
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 519 LQOMOKVWRBSGRK 533
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 IEEAENMORPRKKE 540
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-08-501-572-1
; Sequence 1, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; City: Washington

```

```

STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/501,572
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Tooney, Kimberlin M
REGISTRATION NUMBER: 35,391
REFERENCE/DOCKET NUMBER: 02481.1453-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-501-572-1

```

```

Query Match          24.2%; Score 687.5; DB 3; Length 556;
Best Local Similarity 33.2%; Pred. No. 1.9e-63;
Matches 184; Conservative 88; Mismatches 230; Indels 53; Gaps 11;

```

```

QY 1 MNDYDEVIAFEGWGPFOQLIFELL--SASTIPNGNGMSVFLAGTPEHRCRVPDAA 57
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 1 MPTVDVLEQYGEFGWFOKQAFLLCLISASLAP--IYVGIVLGFPGHVCQNPGEAE 57

QY 58 LSS--AMR-----NNSVP-LRLRDGREVPHSCSRRL-----ATTANFSAL 95
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 58 LSORCGWSQAELNNTYVPELGPDSDEASFLSQCMREYDMNOSTLDCVDPPLSSLVANSOL 117

QY 96 GLEPGRDVLDGLQEESCLDGWFEFSQDYLLSTVTEENMLVCEDMNKVPLTSLFFGVLL 155
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 118 PLGP-----CEHGMY--DTPGSSIVTEFNLYCGDAKMKVDLFGSCVNLGFFL 162

QY 156 GSFVSQGLSDRFRGRKNVLFATMAVOTGFSFLQIFSIEMFTVLYVGMQISNYVAF 215
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 163 GSLVGYIADRRGRKLCILVTTLVTSVSGVLTAVAPDYTSMLLFRLLQGMVSKGSWSGY 222

QY 216 ILGTLLGKSVR---IFSTLGVCTFFAVGYMLPLFAFYTRDMRMMLLATLTPGVLCV 271
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 223 TLTEFVSGSYRTTALIQMA-----FTVGLVGLAGVAVAIIPDMRWLDLAVSLPTFL 277

QY 272 PLWMPFESPRMLISQRRFEADIIQAKAKNNNTAVPAVIFDSVEELNPLKQAKAFILD 331
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 278 LYWVFPEPRMLISQKRTTRAVRIMEQIAQKNKGVPPADLKMCLLEEDASEKRSPPAD 337

QY 332 LEFTRNIAIMTISLLMLTISVGYFALSIDAPNLHGDAVINCFLSALLEIAPYITAML 391
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 338 LEFTPNLRKHTVITLMTLWMSCAVLGYGLIMHGATGANLYLDFEFSLVEFPAAFTILT 397

QY 392 LRTLPRTVITAALFWNGGVLTLFQLVPDYVFLSIGLVMLKFGTISAFSMLYVETAL 451
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 398 IDRIGRITPLAASNIVTGAACLMIFIPHELMHLNTIACIGMGATVILEMVCUNAE 457

QY 452 YPTLVNRMAVGTSTASRSGSIAPYFVY-LGAYNMLPYIYVNGSTLVILGIFTFEPPS 510
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 458 YPTLIRNMGMMVCSALCDLDGIGFTPPMVFRLMEVWQALPLIRGLVGLTAGAMTLLPPT 517

QY 511 LGMFLPETLEOMOKV 525
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 518 KGVALPETIEAENL 532

```

```

RESULT 4
US-09-040-444-1
; Sequence 1, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegans, Henderson, Parabow, Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-040-444-1

```

```

Query Match          24.2%; Score 687.5; DB 3; Length 556;
Best Local Similarity 33.2%; Pred. No. 1.9e-63;
Matches 184; Conservative 88; Mismatches 230; Indels 53; Gaps 11;

```

```

QY 1 MNDYDEVIAFEGWGPFOQLIFELL--SASTIPNGNGMSVFLAGTPEHRCRVPDAA 57
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 1 MPTVDVLEQYGEFGWFOKQAFLLCLISASLAP--IYVGIVLGFPGHVCQNPGEAE 57

QY 58 LSS--AMR-----NNSVP-LRLRDGREVPHSCSRRL-----ATTANFSAL 95
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 58 LSORCGWSQAELNNTYVPELGPDSDEASFLSQCMREYDMNOSTLDCVDPPLSSLVANSOL 117

QY 96 GLEPGRDVLDGLQEESCLDGWFEFSQDYLLSTVTEENMLVCEDMNKVPLTSLFFGVLL 155
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 118 PLGP-----CEHGMY--DTPGSSIVTEFNLYCGDAKMKVDLFGSCVNLGFFL 162

QY 156 GSFVSQGLSDRFRGRKNVLFATMAVOTGFSFLQIFSIEMFTVLYVGMQISNYVAF 215
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 163 GSLVGYIADRRGRKLCILVTTLVTSVSGVLTAVAPDYTSMLLFRLLQGMVSKGSWSGY 222

QY 216 ILGTLLGKSVR---IFSTLGVCTFFAVGYMLPLFAFYTRDMRMMLLATLTPGVLCV 271
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 223 TLTEFVSGSYRTTALIQMA-----FTVGLVGLAGVAVAIIPDMRWLDLAVSLPTFL 277

QY 272 PLWMPFESPRMLISQRRFEADIIQAKAKNNNTAVPAVIFDSVEELNPLKQAKAFILD 331
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 278 LYWVFPEPRMLISQKRTTRAVRIMEQIAQKNKGVPPADLKMCLLEEDASEKRSPPAD 337

```





NUMBER OF SEQUENCES: 17  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: FASTSEQ for Windows Version 2.0b  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/964,127  
 FILING DATE: 06-NOV-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Crews, Ph.D., L. Lee  
 REGISTRATION NUMBER: P-43,567  
 REFERENCE/DOCKET NUMBER: 07334/038001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 TELEEX: 200154  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 520 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: internal  
 US-08-964-127-2

Query Match 9.4%; Score 267; DB 4; Length 520;  
 Best Local Similarity 23.4%; Pred. No. 2.1e-19;  
 Matches 117; Conservative 88; Mismatches 203; Indels 92; Gaps 21.

QY 85 RLATIANFSALGLEPGRVDLGLQESCLDGEFSODVYST-VVTEMNLYCEDNNKVP 143  
 Db 50 RVATSTDPSCSGFAP-----DENHCLKMDYNGLPVLTNAIGMDVCDLGQVY 101  
 QY 144 LTTSLFEVGVLLGSFVSGQLSDRGKKNVLFATMAV-----OTGFSFLQIFS 190  
 Db 102 LEQILFIFGFASGYLFLGYPADRGGRGIVLLTGLVPGCGVGGAAGSSGVMALR--- 158  
 QY 191 ISWEMFTLVFIYVGMGQSNVVAFLIGTELIGKSVRIIFSTLGCYTFEAVGYMLPLFA 250  
 Db 159 -----FLIGFLLAGVD-----LGVYIMRLKLECDPTQRLRVALAGELVGVGHFLGL-A 207  
 QY 251 YFIRDMRLLLATVPGLVCPPLMW--FIPESPRMLISQRRFREADIIOKAANN--- 304  
 Db 208 LVSKDMRFLQRMITAPCLIFLYGWPGLFLESARMLVKQIOEASQVLLKLAERNRPHG 267  
 QY 305 -----NTAVPAVIFDSVEELNPKQOKAF-IIDLFRTNIAIMTMSLLMMLTS----- 353  
 Db 268 QMLGEAQAAL--QDLNCTPLPATSSFSFASLNTYRNI---WNKLLILIGFTNFIHAH 321  
 QY 354 -----VG-----FYALSLDAPNLHGDVYINC-FLSALIEIPAYITALLLRTLPKRY 399  
 Db 322 RHCVQRPVGGGSPSDFLCSLLA---SGTAAACVFLG--VTVDRFGRGICLLLSMTLTG 376  
 QY 400 IIAAVLFWGGGVLLFIQVLPVYVYFELSGVLMGKFGITSAFSLVYFTELYPTLIYRNM 459  
 Db 377 IASIVLL---GLMDYLNEAATITP-----SVIGLFSQAALILSTLLAEVLEPTIYRGR 427  
 QY 460 AVGTSTASRSGSIIP-YEYVIGAVYKRMPLIYVMSGLVLIIGIFLTFEPESIGMTLPET 518  
 Db 428 GLGLIMALGALGGLSGPAORLHMG-HGAFLQHYVLAACALLCILISIMLLPETKRKLIPV 486

QY 519 L---EQMKYKWFERSGKTR 535  
 Db 487 LRDELGRSLRLKQPPTR 506

RESULT 9  
 US-09-496-692-2  
 Sequence 2, Application US/09496692  
 Patent No. 6313271  
 GENERAL INFORMATION:  
 APPLICANT: Grandearl, Andrew David John  
 TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE  
 TITLE OF INVENTION: MOLECULES  
 NUMBER OF SEQUENCES: 17  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: FASTSEQ for Windows Version 2.0b  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/496,692  
 FILING DATE:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/964,127  
 FILING DATE: 06-NOV-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Crews, Ph.D., L. Lee  
 REGISTRATION NUMBER: P-43,567  
 REFERENCE/DOCKET NUMBER: 07334/038001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 TELEEX: 200154  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 520 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: internal  
 US-09-496-692-2

Query Match 9.4%; Score 267; DB 4; Length 520;  
 Best Local Similarity 23.4%; Pred. No. 2.1e-19;  
 Matches 117; Conservative 88; Mismatches 203; Indels 92; Gaps 21.

QY 85 RLATIANFSALGLEPGRVDLGLQESCLDGEFSODVYST-VVTEMNLYCEDNNKVP 143  
 Db 50 RVATSTDPSCSGFAP-----DENHCLKMDYNGLPVLTNAIGMDVCDLGQVY 101  
 QY 144 LTTSLFEVGVLLGSFVSGQLSDRGKKNVLFATMAV-----OTGFSFLQIFS 190  
 Db 102 LEQILFIFGFASGYLFLGYPADRGGRGIVLLTGLVPGCGVGGAAGSSGVMALR--- 158  
 QY 191 ISWEMFTLVFIYVGMGQSNVVAFLIGTELIGKSVRIIFSTLGCYTFEAVGYMLPLFA 250  
 Db 159 -----FLIGFLLAGVD-----LGVYIMRLKLECDPTQRLRVALAGELVGVGHFLGL-A 207  
 QY 251 YFIRDMRLLLATVPGLVCPPLMW--FIPESPRMLISQRRFREADIIOKAANN--- 304  
 Db 208 LVSKDMRFLQRMITAPCLIFLYGWPGLFLESARMLVKQIOEASQVLLKLAERNRPHG 267  
 QY 305 -----NTAVPAVIFDSVEELNPKQOKAF-IIDLFRTNIAIMTMSLLMMLTS----- 353  
 Db 268 QMLGEAQAAL--QDLNCTPLPATSSFSFASLNTYRNI---WNKLLILIGFTNFIHAH 321









QY 424 --FLSIGLVMGKFGITSAFS----MLVYFIAELYPILVRNNAVGVSTASRVGS-IIA 475  
Db 364 MSLYLSIVAI---FGFVAFEEVGGPPIPMFIVALELSOGPRPAIAIVAGFSNMTNFIIVG 419  
QY 476 PFVVLGAVNRMLPYIVNGSLTVLIGIETLF--FPESIGMTLPETLEQOMOKVWFRSGKKT 534  
Db 420 MCFQVBEQCGPYVFIITVLLVLFIRITRYFKYETKGRIFDEI-----ASGFQGGAS 473  
QY 535 RDSMETEE 542  
Db 474 QSDKTPPE 481

## RESULT 15

US-09-031-392-10  
Sequence 10, Application US/0901392  
Patent No. 5942398  
GENERAL INFORMATION:  
APPLICANT: Tartaglia, Louis A.  
APPLICANT: Meng, Xun  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES  
TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/031,392  
FILING DATE: 26-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Melkiohn, Ph.D., Anita L.  
REGISTRATION NUMBER: 35,283  
REFERENCE/DOCKET NUMBER: 07334/072001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 493 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-031-392-10

## Query Match 7.6%; Score 216; DB 2; Length 493;

Best Local Similarity 24.2%; Pred. No. 4.4e-14;  
Matches 111; Conservative 78; Mismatches 171; Indels 98; Gaps 19;

QY 144 LFTSLFVGVVLGSPVSGQLDRGGRKNVLEA--TMAVGTGFSF-LQIFSISEMFTVLF 200  
Db 68 LSVSIFAVGGMIGSLVIGNRKXKAMLVNNVLAIGGLMGLAKKXSFEMILIGR 127  
QY 201 VIVGM-----GQISNTVVAFTLGT-EILGKSVRIIFS-TLGVCTFFAVGYML 245  
Db 128 FIIGLYCGLSSGVPMYGEISPTALRGALGTNLGIVIGILIAOVLGIDSL--LGNS 185  
QY 246 LPLFAFIRDMRMILLATVPGVLCVPLMWFIPSPRWLT-----SQRRRREAE 294  
Db 186 L-----WPLLGLGVPAALQLLLPCCPSPRYLLINKNEAKKALDORLGT 236  
QY 295 DIIOKAKKNNNTAVPAVIFDSVEELNPLKQOKAFIIDLFRTRN-----IAIMTMSLL 348

Db 237 DVSQEVAAEMKD-----ESRKKXSEKXVSVLELFRSRXKQPIYIAIVLQLSOL 285  
QY 349 WMLTVGVYFALSLDAPNLHGDAYLNCPLSALIEIPAYIT-----AMLLRLTL 395  
Db 286 SGINNVFYYSNISI-----PEKAGVGOPYATIGAGVNVPTVVSFVVERA 332  
QY 396 PRR--YIIAAVLFGGCVLLEIQLVPDY--YFLSIGLVMGKFGITSAFS----MLYV 446  
Db 333 GRRTHLHLGLGMAGCAVMTLALALDQVPMMSYVSIVAI--FGFVAFEEVGGPPIPMF 390  
QY 447 FPAELYPILVRNNAVGVSTASRVGS-IIAPYFVLGAVNRMLPYIYNGSLTVLIGIET 504  
Db 391 IVAELFSOGPRPAIAVAGFSNMTNFIIVGLFQYIAELIGPYVFIYFVALLFFIET 450  
QY 505 LFPESIGMTLPETLEQOMOKVWFRSGKTRDSMETEE 542  
Db 451 LKVPETKGRIFDEIAAARKN--KKQPKRESIEELE 486

Search completed: July 17, 2002, 02:03:56  
Job time: 5772 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 01:42:49 ; Search time 58.69 Seconds  
(without alignments)  
902.117 Million cell updates/sec

Title: US-09-521-195B-1

Perfect score: 2845

Sequence: 1 MRDYDEVIAFLGEMGPPQRL.....KTRDSMETENPKVLITRAF 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2218	78.0	557	2	JM0089	organic cation tra
2	2152	75.6	557	2	JEO346	high-affinity carn
3	734.5	26.5	576	2	T22509	hypothetical prote
4	699.5	24.6	593	2	JC4884	organic cation tra
5	681.5	24.0	556	2	S50862	organic cation tra
6	565	19.3	794	2	T27870	hypothetical prote
7	548	19.3	745	2	T16565	hypothetical prote
8	488	17.2	447	2	D89646	protein ZK455.8 [1
9	484.5	17.0	527	2	T01019	transport protein
10	475	16.7	539	2	C96758	probable protein
11	421.5	14.8	518	2	H86299	hypothetical prote
12	412.5	14.5	521	2	H86298	hypothetical prote
13	393	13.8	751	2	C88485	protein F23F12.5 [
14	382.5	13.4	515	2	B96825	hypothetical prote
15	356	12.5	528	2	T21682	hypothetical prote
16	338	11.9	540	2	T25851	hypothetical prote
17	327	11.5	454	2	F75580	probable sugar tra
18	318	11.2	437	2	G97630	probable sugar tra
19	318	11.2	437	2	AB2854	MFS permease [lipo
20	317.5	11.2	422	2	C88504	protein B0361.3 [1
21	317	11.1	455	2	B83213	probable MFS trans
22	312.5	11.0	510	2	B8381	protein T22F.1 [1
23	312	11.0	450	2	F95360	probable transmem
24	311	10.9	591	2	T30895	sugar transport pr
25	307	10.8	435	2	T15290	hypothetical prote
26	305	10.7	461	2	AB3208	MFS permease [lipo
27	294.5	10.4	529	2	T23190	hypothetical prote
28	293.5	10.3	452	2	AD0300	probable transport
29	290	10.2	400	2	C69757	transporter homoio

30	290	10.2	448	2	AC3602	transporter, mfs s
31	286.5	10.1	478	2	T33985	hypothetical prote
32	284.5	10.0	648	2	E88960	protein T08B1.1 [1
33	281	9.9	422	2	G72234	hypothetical prote
34	276	9.7	461	2	D70073	metabolic transpo
35	275	9.7	520	2	T23545	hypothetical prote
36	274.5	9.6	443	2	E64725	yeau protein - Esc
37	274.5	9.6	443	2	H90634	probable transport
38	274.5	9.6	442	2	H85485	probable transport
39	272.5	9.6	442	2	A83122	probable MFS trans
40	271.5	9.5	524	2	T27082	hypothetical prote
41	271	9.5	423	2	T19030	hypothetical prote
42	269.5	9.5	423	2	S74046	hypothetical sugar tra
43	262	9.2	469	2	G65058	hypothetical prote
44	259.5	9.1	480	2	T23608	hypothetical prote
45	248	8.7	541	2	T28069	hypothetical prote

ALIGNMENTS

Query Match	78.0%; Score 2218; DB 2; Length 557;
Best Local Similarity	75.9%; Pred. No. 1.7e-155;
Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;	
QY	1 MRDYDEVIAFLGEMGPPQRLIFFLSASIPNGFNGSVFLAGPEHRCVPDAANLSS 60
DB	1 MRDYDEVIAFLGEMGPPQRLIFFLSASIPNGFNGSVFLAGPEHRCVPDAANLSS 60
QY	61 AMRNNSVPLRLRDGREVPHSCSRYLATTANFSALGLEPGRDVLDGLEQESCLDGWERS 120
DB	61 AMRNHTVPLRLRDGREVPHSCSRYLATTANFSALGLEPGRDVLDGLEQESCLDGWERS 120
QY	121 QDVLSTIVTEEMNLVCCDNKKVPLTSLFVGVGLGSFVSGQISDPRGRNVLFAFMAVO 180
DB	121 QDVLSTIVTEEMNLVCCDNKKVPLTSLFVGVGLGSFVSGQISDPRGRNVLFAFMAVO 180
QY	181 TGFSFLDIFSTSMEMFTVLVFIVMGQISNYVAFIIGTEIILKSVRIIFSTLGVCTFFA 240
DB	181 TGFSFLDIFSTSMEMFTVLVFIVMGQISNYVAFAVGTETILKSVRIIFSTLGVCTFFA 240
QY	241 VGYMLPLFAVFIKDMMLLALTVPGVLCVPLMFTIPESPRMLISQRRPREADIIQKA 300
DB	241 VGYMLPLFAVFIKDMMLLALTVPGVLCVPLMFTIPESPRMLISQRRPREADIIQKA 300
QY	301 AKMNTVAVPIVD--SVEELNPLKOOKAPLILDFRRNTAIMTIVISLIMLTSGYFA 358
DB	301 AKAGIVVPSTIPDSSELQDLSKKQSHNILLDKRTWNTIMVITISVGYFG 360
QY	359 LSLDAPNLHGDAYLNCFLSLALIEIPAYITAMLLILRTLPRIYITAAVLFMGVGLFIOY 418
DB	361 LSLDTPNLHGDIFVNCFLSAMVEPAVAVLAMLLOQYIPRFRYSNATLALFGSGVLLFMQY 420
QY	419 PVDYFPLSLGLVLMGKFGITSAFSMLYVFTAEILYPTLVRRMAVGVITSAKVGSIITAPYF 478

[illegible]

Query Match	75.6%;	Score 2152;	DB 2;	Length 557;
Best Local Similarly	73.1%;	Pred. No. 1.2e-150;		
Matches 407;	Conservative 68;	Mismatches 76;	Indels 6;	Gaps 3;

```

QY 1 MRDDEVAFEGEGBPQFORLFEFLLSAIIIPNGNGSVVLAETPEHRCPVPAANLSS 60
Db 1 MRDDEVAFEGEGBPQFORLFEFLLSAIIIPNGNGSVVLAETPEHRCPVPAANLSS 60
QY 61 AMRNNSVPLRLRDGREVPHSCSRYRLATIANFSALGIEPRHDVLDGOLDESCJDGEFS 120
Db 61 AMRNNSIPLFETKDRQVPOSCRRYRLATIANFSALGIEPRHDVLDGOLDEONCJDGMEYN 120
QY 121 ODVLYSTVYTRMNLVCEBENMKVPLTTSLEFYGVLLGSFVSQSLSDRGRKNVLPATMAVQ 180
Db 121 ODVLYSTVYTRMNLVCEBENMKVPLTTSLEFYGVLLGSFVSQSLSDRGRKNVLPATMAVQ 180
QY 181 TGSEFLOLFEISMEFNTLFEYVGMGOISNVVAFIITELGKSVRIITESTJGVCFFA 240
Db 181 TGSEFLOLFEISMEFNTLFEYVGMGOISNVVAFIITELGKSVRIITESTJGVCFFA 240
QY 241 VGYMLPLFAFTRIDWRMLLALTVPGLVCPLMFTIPESBRLMISORPREADIIOKA 300
Db 241 VGYMLPLFAFTRIDWRMLLALTVPGLVCPLMFTIPESBRLMISORPREADIIOKA 300
QY 301 AKMNTAVPAVIFD--SYEELNPLKOOKAFILDLFRIRNIALMTINSLLIMLTSVGYFA 358
Db 301 AKMNTAVPAVIFD--SYEELNPLKOOKAFILDLFRIRNIALMTINSLLIMLTSVGYFA 358
QY 359 LSTDAPNHLHGAVYNCLSLALIEIPAYITAMLLRLTPRBYIAAIVFMGGVILFQIV 418
Db 359 LSTDAPNHLHGAVYNCLSLALIEIPAYITAMLLRLTPRBYIAAIVFMGGVILFQIV 418
QY 419 PVDYFPLSIGLVMLGKSGITSASFMLVFTAEIYPTLVRRMNAVGVTSTASRGSIAPYE 478
Db 419 PVDYFPLSIGLVMLGKSGITSASFMLVFTAEIYPTLVRRMNAVGVTSTASRGSIAPYE 478
QY 479 VYLGAVYRMPLPYIVMGSJLVJLIGITFLFEPESJGMLTPETLEOMOKVWERSK--KTR 535
Db 479 VYLGAVYRMPLPYIVMGSJLVJLIGITFLFEPESJGMLTPETLEOMOKVWERSK--KTR 535
QY 536 DSMETEENPKVL--ITAF 551
Db 536 DSMETEENPKVL--ITAF 551
QY 551 TQKDGESPTVLKSTAR 557
Db 551 TQKDGESPTVLKSTAR 557

```

RESULT 3  
T22509  
hypotheical protein F52F12.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22509  
R:Matthews, L.  
submitted to the EMBL Data Library, December 1996  
A:Reference number: Z19573  
A:Accession: T22509  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-576 <WIL>  
A:Cross-references: EMBL:Z83228; PIDN:CAB05732.1; GSPDB:GN00019; CESP:F52F12.1  
A:Experimental source: clone F52F12  
C:Genetics:  
A:Map: CESP:F52F12.1  
A:Map position: 1  
A:Introns: 16/1; 23/3; 50/1; 80/3; 108/2; 134/3; 221/1; 257/2; 319/3; 358/2; 4

Query Match	26.5%	Score 754.5	DB 2	Length 576
Best Local Similarity	30.2%	Pred No. 7.2e-48		
Matches 169	Conservative 119	Mismatches 236	Indels 35	Gaps 9

OY	3	DYEVIAITLBSGMPQRIJLIEFLLS-ASTITPBNFGMSVFLAGIPBHRICVPAIDANLSSA	61
Dd	16	DFDFLEOVANGYQIVAEFFELICLPLTSPASFAFAMNIPFVGNBPHTCHPECKEYLRP	75
OY	62	WRNNSVPLRLDGRGEVPHSCSRRLATITANSALGEPBGRVDLGOJESQSCDGEFSQ	121
Dd	76	LINDQIL-----SCKQYMETQINFRATTSAP-VDITYSDRLSLPCQMGNDYD	124
OY	122	DVYLSVTVENNLVCEDNMKVPLTTSLEFVGVLGGSVQSGOLDRFGKVLVEATMAVOT	181
Dd	125	STYLDSTLVEENLVCDOOAMIEISTSFYSGFIGNCLFGVADKFGFRSFPVILTVLI	184
OY	182	GFSELOJFSISEMEFTYLFVIVGGOISNVYVNIITGTEILGKSVRTIESTLGCTFFAV	241
Dd	185	VCGTASSFAKIOESFTIITREFTGLAEPALQIPIITICMEFGNSGR-IFGGLMTSFFGA	243
OY	242	GVMLLPFEAYFIRIMRMILLATYPGVLCVPLMFIPESPRMLISQRPFEADITOKAA	301
Dd	244	AMALLGVAMHIRMRQLOTFPCNAPFAFIITYYFLEPSRMSVSGKADAKQOLKKA	303
OY	302	KAN--NTAPAYFDSVEEINPLKOOKAF----IIDEFTRNIAIMTJMSLLMLTIS	353
Dd	304	KMNKSNVDVDELY-DSMKHQNQNAEKEKTRKSHNVYDEKTEPNLRKKTILVYIWMNA	362
OY	354	VGYFALSIDAPNLDGADYINCFALSALIEPAYITAWILLETTLRRYITAAVLFWGGVLL	413
Dd	363	IITGGLTINLSNLEVDYDWSITINGAELDEGFEYVWMLLQACGRMMLTATMLVCSIGCV	422
OY	414	FIOLEPVVDYFELSIGLVMLGKFGTTSASFMSLVEYETALETYTLVRNNAVGTSPTASRGSI	473
Dd	423	SAMNPDGYPMLVASAFICFGYGSFAYIITIFAGLITTYVVARAIGMSSNVAASGLL	482
OY	474	IAPFYVLGAYNRMLPYIVMGSLTVLIGITFLEPPESLGMTLPELTLEOMOKVYKFRSGKK	533
Dd	483	LAPIVMLGKIVKILPLITLIMGMLMSAGILTFLEPBTLAGAPLMTIEDAENF-----GKK	537
OY	534	-----TRDSMETEENP	544
Dd	538	PEPDGSMFTQAAKKRESOP	556

RESULT 4  
JC4884 organic cation transporter protein 2 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Sep-1996 #sequence\_revision 18-Oct-1996 #text\_change 05-Nov-1999  
#Accession: JC4884



R:Okuda, M.; Saito, H.; Urakami, Y.; Takano, M.; Inui, K.  
 Biochem. Biophys. Res. Commun. 224, 500-507, 1996  
 A:Title: cDNA cloning and functional expression of a novel rat kidney organic cation transporter  
 A:Reference number: J04884; MUID:96295517  
 A:Accession: J04884  
 A:Molecule type: mRNA  
 A:Residues: 1-593 <OK>  
 A:Cross-references: DDBJ:D83044; NID:91502282; PIDN:BA11754.1; PID:dl012421; PID:915022  
 A:Experimental source: Kidney  
 C:Comment: This protein is responsible for the transport of cationic drugs in kidney.

Query Match 24.6%; Score 699.5; DB 2; Length 593;  
 Best Local Similarity 34.0%; Pred. No. 8.3e-44;  
 Matches 189; Conservative 87; Mismatches 243; Indels 37; Gaps 14;

QY 1 MRDYDEVIAFLGEMGPORLIFELLASITPENGNG--MSVYFLAGTPERHCRVPDANL 58  
 1 MSYVDLILHIEHFEHFKQFFELL--ALLSAFPIPIVIGVIFLPTPDHHCMSPGAAL 58  
 DB 1 MSYVDLILHIEHFEHFKQFFELL--ALLSAFPIPIVIGVIFLPTPDHHCMSPGAAL 58  
 QY 59 SS--AMR-----NSYV-LRLDGRVPHSCRRY-----LATIANFSALGLEPGADV 103  
 59 SORCGMSOAEELNYTPVPGPSDEASFLSOCRKYEVDMNOSTLDCVPLSSIA-----A 112  
 DB 59 SORCGMSOAEELNYTPVPGPSDEASFLSOCRKYEVDMNOSTLDCVPLSSIA-----A 112  
 QY 104 DLGOLROESCLDGMFESODVYLTSTVTENMLVCEEDNMKVPITTSLEFVGLSGFSQOL 163  
 113 DRKQPLRGCEHGWYNTG--SSIVTERNLVCAHSMLEDFQSYVNVNGFICAMMIGYL 170  
 DB 113 DRKQPLRGCEHGWYNTG--SSIVTERNLVCAHSMLEDFQSYVNVNGFICAMMIGYL 170  
 QY 164 SDRFGRKNVLFAMAVQGFSLQIFSLISEMFTLVLEVIGQISNVYAFILGTEILG 223  
 171 ADRFGKFCFLVYILINASIGALMAISPYAMLVFRLOGLVSKAGWLIGYILITEFVG 230  
 DB 171 ADRFGKFCFLVYILINASIGALMAISPYAMLVFRLOGLVSKAGWLIGYILITEFVG 230  
 QY 224 KSVRIFFSLGVC--TFEAVGYMLPLFAFYFRDMRMILLALTGVGLCVLMTPIESP 281  
 231 LGYR--RWVGLCYQIAFVGLITLAGVAYVLPNMRLQFAVTLNFCFCFLYFMCITPSP 287  
 DB 231 LGYR--RWVGLCYQIAFVGLITLAGVAYVLPNMRLQFAVTLNFCFCFLYFMCITPSP 287  
 QY 282 RMLISORRREAEEDIIQKAKMNTAVPAVIFDSVEELNPKQOKAFITLDFRTNIAIM 341  
 288 RMLISQNKIVKAMKIKIKHAKKNGKSPVSLQNLTPDDAGKIKLPISLIDVTRPQIKH 347  
 DB 288 RMLISQNKIVKAMKIKIKHAKKNGKSPVSLQNLTPDDAGKIKLPISLIDVTRPQIKH 347  
 QY 342 TMSLLMLTSGYFALSIDAPNLHGD--AYINCLSLALIEIPATITMLLRLPRXYI 400  
 348 TILMLNMTSSVLYQGLIMEM--GLAGDNIYDFEYSALVERPAFIIILIDRGRXP 406  
 DB 348 TILMLNMTSSVLYQGLIMEM--GLAGDNIYDFEYSALVERPAFIIILIDRGRXP 406  
 QY 401 IAAVLWGGGVLLFIQLVVDYFLSIGLVMKFGITSAPSEMLVFAEELPYLVRMA 460  
 407 WAYSNNVAGAACLASVFIIDDLQWLKITTAGCGRMGITMAYEMVGLVAAELPYLRNLG 466  
 DB 407 WAYSNNVAGAACLASVFIIDDLQWLKITTAGCGRMGITMAYEMVGLVAAELPYLRNLG 466  
 QY 461 VGTSTASRVSIIAPFYV--LGATNRMLPYIVMGLVYLIGIFLFPESIGMTPELT 519  
 467 VLVCSMCDIGITLPELVYRLTIMERPVLVFAVGLVAGALVLLPETGKALPETI 526  
 DB 467 VLVCSMCDIGITLPELVYRLTIMERPVLVFAVGLVAGALVLLPETGKALPETI 526  
 QY 520 EQMKVAKFRSGKTR 535  
 527 EDAENMQ--RPRKKR 540  
 DB 527 EDAENMQ--RPRKKR 540

RESULT 5  
 S0862  
 Organic cation transporter protein OCT1 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 05-Nov-1999  
 C:Accession: S50862; S78533; I58089  
 R:Guendemann, D.; Gorboulev, V.; Gambaryan, S.; Veyhl, M.; Koepsell, H.  
 Nature 372, 549-552, 1994  
 A:Title: Drug excretion mediated by a new prototype of polyspecific transporter.  
 A:Reference number: I58089; MUID:95082907  
 A:Accession: S50862  
 A:Molecule type: mRNA  
 A:Residues: 1-556 <CRU>  
 A:Cross-references: EMBL:X78655; NID:9633621; PIDN:CA55411.1; PID:9633622  
 A:Experimental source: Kidney  
 R:Gorboulev, V.G.

submitted to the EMBL Data Library, January 1995  
 A:Reference number: S78533  
 A:Accession: S78533  
 A:Molecule type: mRNA  
 A:Residues: 1-342, 'N', 344-556 <GOR>  
 A:Cross-references: EMBL:X78655; NID:9633621; PIDN:CA55411.1; PID:9633622  
 C:Keywords: glycoprotein; phosphoprotein; transmembrane protein  
 F:20-46/Domain: transmembrane #status predicted <TM1>  
 F:154-171/Domain: transmembrane #status predicted <TM2>  
 F:178-197/Domain: transmembrane #status predicted <TM3>  
 F:243-260/Domain: transmembrane #status predicted <TM4>  
 F:267-283/Domain: transmembrane #status predicted <TM5>  
 F:350-366/Domain: transmembrane #status predicted <TM6>  
 F:380-398/Domain: transmembrane #status predicted <TM7>  
 F:406-425/Domain: transmembrane #status predicted <TM8>  
 F:435-452/Domain: transmembrane #status predicted <TM9>  
 F:469-485/Domain: transmembrane #status predicted <TM10>  
 F:494-514/Domain: transmembrane #status predicted <TM11>  
 F:127-137/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pred  
 F:186/292/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status pred  
 F:296/343/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status pred

Query Match 24.0%; Score 681.5; DB 2; Length 556;  
 Best Local Similarity 33.0%; Pred. No. 1.6e-42;  
 Matches 183; Conservative 88; Mismatches 231; Indels 53; Gaps 11;

QY 1 MRDYDEVIAFLGEMGPORLIFELL--SASTIENGNGMSVYFLAGTPERHCRVPDANL 57  
 1 MPTVDVILHIEHFEHFKQFFELL--ALLSAFPIPIVIGVIFLPTPDHHCMSPGAAL 57  
 DB 1 MPTVDVILHIEHFEHFKQFFELL--ALLSAFPIPIVIGVIFLPTPDHHCMSPGAAL 57  
 QY 58 LSS--AMR-----NSYV-LRLDGRVPHSCRRY-----LATIANFSAL 95  
 58 SORCGMSOAEELNYTPVPGPSDEASFLSOCRKYEVDMNOSTLDCVPLSSIA-----A 112  
 DB 58 SORCGMSOAEELNYTPVPGPSDEASFLSOCRKYEVDMNOSTLDCVPLSSIA-----A 112  
 QY 58 LSS--AMR-----NSYV-LRLDGRVPHSCRRY-----LATIANFSAL 95  
 58 LSGRCMSOAEELNYTPVPGPSDEASFLSOCRKYEVDMNOSTLDCVPLSSIA-----A 112  
 DB 58 LSGRCMSOAEELNYTPVPGPSDEASFLSOCRKYEVDMNOSTLDCVPLSSIA-----A 112  
 QY 96 GLEPGADVLOGLROESCLDGMFESODVYLTSTVTENMLVCEEDNMKVPITTSLEFVGL 155  
 118 PLGP-----CEHGWY--DRPGSSIVTERNLVCAHSMLEDFQSYVNVNGFICAMMIGYL 162  
 DB 118 PLGP-----CEHGWY--DRPGSSIVTERNLVCAHSMLEDFQSYVNVNGFICAMMIGYL 162  
 QY 156 GSFVSGQLSDRGKKNVLFAMAVQGFSLQIFSLISEMFTLVLEVIGQISNVYAFILGTEILG 215  
 163 GSVLVGYIADRGKKNVLFAMAVQGFSLQIFSLISEMFTLVLEVIGQISNVYAFILGTEILG 215  
 DB 163 GSVLVGYIADRGKKNVLFAMAVQGFSLQIFSLISEMFTLVLEVIGQISNVYAFILGTEILG 215  
 QY 216 ILGTEILGKSVR---IISTGLVCTFEAVGYMLPLFAFYFRDMRMILLALTGVGLCV 271  
 223 TLTEFVSGSYRRTATLVQMA-----FTVGLVGLAGVAYVLPNMRLQFAVTLNFCFCFLYFMCITPSP 277  
 DB 223 TLTEFVSGSYRRTATLVQMA-----FTVGLVGLAGVAYVLPNMRLQFAVTLNFCFCFLYFMCITPSP 277  
 QY 272 PLWMTFPESEPRMLISORRREAEEDIIQKAKMNTAVPAVIFDSVEELNPKQOKAFITLDFRTNIAIM 331  
 278 LYTWEVPESEPRMLISORRREAEEDIIQKAKMNTAVPAVIFDSVEELNPKQOKAFITLDFRTNIAIM 331  
 DB 278 LYTWEVPESEPRMLISORRREAEEDIIQKAKMNTAVPAVIFDSVEELNPKQOKAFITLDFRTNIAIM 331  
 QY 332 LFTRTNIAIMTMSLLMLTSGYFALSIDAPNLHGDAYINCLSLALIEIPATITMLLRLPRXYI 391  
 338 LFTPTPLRKHTVILMYLWMSCAVLQGLIMHGAGANLYIDFEYSALVERPAFIIILVTRPQIKH 397  
 DB 338 LFTPTPLRKHTVILMYLWMSCAVLQGLIMHGAGANLYIDFEYSALVERPAFIIILVTRPQIKH 397  
 QY 392 LRTPLPRXYIILAVLFWGGVLLFIQLVVDYFLSIGLVMKFGITSAPSEMLVFAEELPYLVRMA 451  
 398 IDRIIGRTPIAASNLVTGAACLMIFIPHELMNLVTLTACGRMAATVILDMVCLVNL 457  
 DB 398 IDRIIGRTPIAASNLVTGAACLMIFIPHELMNLVTLTACGRMAATVILDMVCLVNL 457  
 QY 452 YPFLVNRMAVGTSTASRVSIIAPFYV--LGATNRMLPYIVMGLVYLIGIFLFPESIGMTPELT 510  
 458 YPFLVNRMAVGTSTASRVSIIAPFYV--LGATNRMLPYIVMGLVYLIGIFLFPESIGMTPELT 510  
 DB 458 YPFLVNRMAVGTSTASRVSIIAPFYV--LGATNRMLPYIVMGLVYLIGIFLFPESIGMTPELT 510  
 QY 511 LGMTLPETLEQOKV 525  
 518 KVALPETIEEENL 532  
 DB 518 KVALPETIEEENL 532

RESULT 6  
 T27870  
 hypothetical protein ZK455.8 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000









C:Accession: T21682  
R:Steward, C.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: 219458  
A:Accession: T21682  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-528 <Wild>  
A:Cross-references: EMBL:Z81524; PIDN:CAB04250.1; GSPDB:GN00023; CESP:F32H5.4  
A:Experimental source: clone F32H5  
C:Genetics:  
A:Gene: CESP:F32H5.4  
A:Map position: 5  
A:introns: 63/1; 108/3; 148/3; 249/2; 408/1; 472/3

Query Match 12.5%; Score 356; DB 2; Length 528;  
Best Local Similarity 25.9%; Pred. No. 1.3e-18;  
Matches 125; Conservative 86; Mismatches 174; Indels 98; Gaps 21;

```
QY 100 GHDVDLGLQDESCLDGMEF-----SQDYLTSTVYTEMNLVCEDNMKVPLTSLF 149
      | | | : : : | : | : | : : : : : | : | : : : | : | :
DB 73 GPDVFI---KDICKSSDFDNPNCTITSKYDF--SINVDYGHFCGAWVKTSISVQ 126

QY 150 FVGVLGSGFVSQGLSDRFGRKNVLFATMAVOTGFSEFLOFISWEMFTVLFYIGM---G 206
      ||||:| | :||:| | | : : : | | : | : | : | : | : | : |
DB 127 MGVGLIGSVTSGAVADRGRLKVLSCFEMVSSLSILNTPAKDLIFTIIRILSIFKGG 186

QY 207 QISNIVVAFIIGTELKSVRIISTLGCTFFPAVGIMLPLFAFIRDW---RMLLAL 263
      : | | : : : | : | : : | : | : | : | : | : | : | : |
DB 187 LLSTGV---YKMEHPVPRQHRFWIATW---ISWAPNMILSFVAYLCHDMITYOYAIFAL 240

QY 264 TVPGVLCVPLMWFIPESPRLISQRPFEADIIORAKKN-NTAVPAVIFDSVEEL--- 319
      : || : : | : ||||| | | | : : : || : : : || : : : || :
DB 241 SMRGAV---VLFVKESPRMLIQAGRIEAKOYLKHHIMEVDGNTSDHS--WNEIEMLOT 295

QY 320 -----NFLKQOKAF-IIDLFRTNIAIMTMSLLMM--LTSVGFALSLDAPNLHGD 369
      : : : | : : | : : | : : | : : | : : | : : | : : |
DB 296 EKKROEERIKRKNYDFRHLFWMNKYMASVT--MILMLGMFSTSTFNNGVFENIEKLSGS 352

QY 370 ATLNCFLSALIEIPAYITAMLLRTLPRIYIAAVLFWGGVLLFTOL----- 417
      | : | : : | : | : | : | : | : | : | : | : | : | : |
DB 353 LYINALLMGSUR-----WVL-----NIFPGIADLKPAKLGKRHHILSKLT 393

QY 418 -----VPDYIF-----LSIGL-VMLGKFGITSAPSMLYFTAEIYPTLVNMAVG 462
      : | | | : : | | | : : | : : | : : | : : | : : | : |
DB 394 ITICVFSIFVTYYEYEDYSILIRVATLLASATASQVFITKSMVLMFEYPTVIRNSAVS 453

QY 463 VTSTASRGSIIPAFVYVLGAYNMLPYIYWSLTVLIGI-FTLFPESIGMTLPETLEO 521
      | : |||:| : | : | : || : | : | : | : | : | : | : | :
DB 454 FKSSASRGITIGLPOLFICPY-KSLPYALITGFCLEDAIARQLRPEITKGRPLPETMPE 512

QY 522 MOK 524
      |
DB 513 RHK 515
```

Search completed: July 17, 2002, 02:05:09  
Job time: 1340 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 02:03:59 ; Search time 39.73 Seconds  
(without alignments)

536,986 Million cell updates/sec

Title: US-09-521-195b-1

Perfect score: 2845  
Sequence: 1 MRDYEVIAFLGKWPQRL.....KTRDSMETENPKVLITAF 551

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2218	78.0	557	1	OCN2_HUMAN
2	2165	76.1	557	1	OCN2_RAT
3	2164	76.1	557	1	OCN2_MOUSE
4	393	13.8	751	1	YLN3_CAEEL
5	317.5	11.2	1222	1	YMP3_CAEEL
6	307	10.8	435	1	YTI3_CAEEL
7	294.5	10.4	529	1	YQUL_CAEEL
8	290	10.2	400	1	YCEI_CAEEL
9	274.5	9.6	443	1	YAAD_ECOLI
10	270	9.5	461	1	YASB_CAEEL
11	262	9.2	445	1	YCGS_ECOLI
12	248	8.7	490	1	GTRI_CHICK
13	247	8.7	592	1	HXT5_YEAST
14	245	8.6	559	1	YDJR_ECOLI
15	245	8.6	566	1	KHT2_KLUFA
16	244.5	8.6	567	1	HXT9_YEAST
17	243	8.5	451	1	GTRI_PIG
18	243	8.5	492	1	GTRI_RAT
19	240	8.4	493	1	GTRI_CANFA
20	238	8.4	492	1	GTRI_BOVIN
21	238	8.4	492	1	GTRI_MOUSE
22	237.5	8.3	452	1	YDJR_ECOLI
23	237	8.3	492	1	GTRI_HUMAN
24	237	8.3	546	1	HXT0_YEAST
25	237	8.3	567	1	HXTA_YEAST
26	236	8.3	451	1	YTAJ_BACSU
27	234.5	8.2	494	1	GTRI_SHEEP
28	234	8.2	491	1	YXLE_ECOLI
29	231	8.1	522	1	YXLE_ECOLI
30	230	8.1	494	1	GTRI_BOVIN
31	226	7.9	457	1	YIRO_YEAST
32	225	7.9	492	1	GTRI_RABIT
33	225	7.9	763	1	RGT2_YEAST

34	223	7.8	413	1	MUCK_ACICA	P94131 acinetobact
35	222	7.8	742	1	SVY2_RAT	O02563 rattus norv
36	221.5	7.8	570	1	HXT6_YEAST	P39003 saccharomyc
37	221.5	7.8	570	1	HXT7_YEAST	P39004 saccharomyc
38	221	7.8	523	1	GTRI_MOUSE	P14246 mus musculu
39	220	7.7	496	1	GTRI_HUMAN	P11169 homo sapien
40	219.5	7.7	472	1	ARAB_KLEOX	P45598 klebsiella
41	218	7.7	569	1	HXT8_YEAST	P40866 saccharomyc
42	216.5	7.6	466	1	BENK_ACICA	O30513 acinetobact
43	215.5	7.6	472	1	ARAB_ECOLI	P09830 escherichia
44	215.5	7.6	551	1	HGT1_KLUFA	P49374 kluyveromyc
45	215	7.6	448	1	PCAK_PSEPU	O51955 pseudomonas

## ALIGNMENTS

Result	ID	OCN2_HUMAN	STANDARD	PRT	557 AA
1	OCN2_HUMAN	076082	16-OCT-2001 (Rel. 40, Created)		
AC	OCN2_HUMAN	076082	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	OCN2_HUMAN	076082	16-OCT-2001 (Rel. 40, Last annotation update)		
DR	OCN2_HUMAN	076082	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	OCN2_HUMAN	076082	Organic cation/carnitine transporter 2 (solute carrier family 22, member 5) (High-affinity sodium-dependent carnitine cotransporter).		
GN	OCN2_HUMAN	076082	SLC22A5 OR OCTN2.		
OS	OCN2_HUMAN	076082	Homo sapiens (Human).		
OC	OCN2_HUMAN	076082	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OX	OCN2_HUMAN	076082	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI	OCN2_HUMAN	076082	NCBI_TaxID=9606;		
11	OCN2_HUMAN	076082	SEQUENCE FROM N.A.		
RP	OCN2_HUMAN	076082	SEQUENCE FROM N.A.		
RX	OCN2_HUMAN	076082	MEDLINE=98289574; PubMed=9618255;		
RA	OCN2_HUMAN	076082	Wu X., Prasad P.D., Leibach F.H., Ganapathy V.;		
RT	OCN2_HUMAN	076082	"CDNA sequence, transport function, and genomic organization of human		
RT	OCN2_HUMAN	076082	OCTN2, a new member of the organic cation transporter family.";		
RL	OCN2_HUMAN	076082	Biochem. Biophys. Res. Commun. 246:589-595(1998).		
12	OCN2_HUMAN	076082	SEQUENCE FROM N.A.		
RP	OCN2_HUMAN	076082	SEQUENCE FROM N.A.		
RX	OCN2_HUMAN	076082	MEDLINE=98352077; PubMed=9685390;		
RA	OCN2_HUMAN	076082	Tamai I., Ohashi R., Nezu J.-I., Yabuuchi H., Oku A., Shimane M.,		
RT	OCN2_HUMAN	076082	Sai Y., Tsuji A.;		
RT	OCN2_HUMAN	076082	"Molecular and functional identification of sodium ion-dependent, high		
RL	OCN2_HUMAN	076082	affinity human carnitine transporter OCTN2.";		
13	OCN2_HUMAN	076082	J. Biol. Chem. 273:20378-20382(1998).		
14	OCN2_HUMAN	076082	SEQUENCE FROM N.A.		
RP	OCN2_HUMAN	076082	SEQUENCE FROM N.A.		
RX	OCN2_HUMAN	076082	MEDLINE=99113835; PubMed=9916797;		
RA	OCN2_HUMAN	076082	Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,		
RT	OCN2_HUMAN	076082	Nakido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsushita T.,		
RA	OCN2_HUMAN	076082	Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,		
RT	OCN2_HUMAN	076082	Tsuji A.;		
RT	OCN2_HUMAN	076082	"Primary systemic carnitine deficiency is caused by mutations in a		
RT	OCN2_HUMAN	076082	gene encoding sodium ion-dependent carnitine transporter.";		
RL	OCN2_HUMAN	076082	Nat. Genet. 21:91-94(1999).		
141	OCN2_HUMAN	076082	CHARACTERIZATION.		
RP	OCN2_HUMAN	076082	CHARACTERIZATION.		
RX	OCN2_HUMAN	076082	MEDLINE=99384224; PubMed=10454528;		
RA	OCN2_HUMAN	076082	Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,		
RT	OCN2_HUMAN	076082	Chen J., Conway S.J., Ganapathy V.;		
RT	OCN2_HUMAN	076082	"Functional characteristics and tissue distribution pattern of organic		
RT	OCN2_HUMAN	076082	cation transporter 2 (OCTN2), an organic cation/carnitine		
RT	OCN2_HUMAN	076082	transporter.";		
RL	OCN2_HUMAN	076082	J. Pharmacol. Exp. Ther. 290:1462-1492(1999).		
151	OCN2_HUMAN	076082	VARIANT CDSP GLN-169.		
RP	OCN2_HUMAN	076082	VARIANT CDSP GLN-169.		
RX	OCN2_HUMAN	076082	MEDLINE=99355597; PubMed=10425211;		
RA	OCN2_HUMAN	076082	Burwinkel B., Kreuder J., Schweitzer S., Vorgerd M., Gempel K.,		
RT	OCN2_HUMAN	076082	Gerditz K.-D., Kiliann M.W.;		
RT	OCN2_HUMAN	076082	"Carnitine transporter OCTN2 mutations in systemic primary carnitine		
RT	OCN2_HUMAN	076082	deficiency: a novel Arg169Gln mutation and a recurrent Arg282ter		







DT 16-OCT-2001 (Rel. 40, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Organic cation/carnitine transporter 2 (Solute carrier family 22,  
 DE member 5) (High-affinity sodium-dependent carnitine cotransporter).  
 GN SLC22A5 OR OCTN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=99113835; PubMed=9916797;  
 RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,  
 RA Nishimoto H., Sai Y., Kojima A., Shoji Y., Takada G., Matsushita T.,  
 RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,  
 RA Tsuji A.;  
 RT "Primary systemic carnitine deficiency is caused by mutations in a  
 RT gene encoding sodium ion-dependent carnitine transporter.";  
 RL Nat. Genet. 21:91-94(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND VARIANT JVS ARG-352.  
 RC STRAIN=C3H;  
 RX MEDLINE=99057546; PubMed=9837751;  
 RA Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;  
 RT "A missense mutation of mouse OCTN2, a sodium-dependent carnitine  
 RT cotransporter, in the juvenile visceral steatosis mouse.";  
 RL Biochem. Biophys. Res. Commun. 252:590-594(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
 RC STRAIN=C3H;  
 RX MEDLINE=99384224; PubMed=10454528;  
 RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,  
 RA Chen J., Conway S.J., Ganapathy V.;  
 RT "Functional characteristics and tissue distribution pattern of organic  
 RT cation transporter 2 (OCTN2), an organic cation/carnitine  
 RT transporter";  
 RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).  
 CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE  
 CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE  
 CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF  
 CC CARNITINE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL  
 CC STEATOSIS (JVS).  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
 CC CATION SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB015800; BAA36590.1; -  
 CC EMBL: AF111425; AAC99787.1; -  
 CC EMBL: AF10417; AAD54060.1; -  
 CC MGI: MGI:1329012; Sicz22a5.  
 DR InterPro: IPR003662; sub-transporter.  
 DR Pfam: PF00083; sugar\_tr\_1.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
 KW Transport; Transmembrane; Glycoprotein; Disease mutation.  
 FT TRANSMEM 21 41 POTENTIAL.  
 FT TRANSMEM 143 163 POTENTIAL.  
 FT TRANSMEM 173 193 POTENTIAL.  
 FT TRANSMEM 198 218 POTENTIAL.  
 FT TRANSMEM 233 253 POTENTIAL.  
 FT TRANSMEM 258 278 POTENTIAL.  
 FT TRANSMEM 342 362 POTENTIAL.  
 FT TRANSMEM 374 394 POTENTIAL.  
 FT TRANSMEM 407 427 POTENTIAL.  
 FT TRANSMEM 431 451 POTENTIAL.

FT TRANSMEM 489 509 POTENTIAL.  
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 352 352 L -> R (IN JVS).  
 SQ SEQUENCE 557 AA; 62779 MW; 6093F0BE9612B204 CRC64;  
 Query Match 76.1%; Score 2164; DB 1; Length 557;  
 Best Local Similarity 72.9%; Pred. No. 7.7e-133;  
 Matches 406; Conservative 70; Mismatches 75; Indels 6; Gaps 3;  
 QY 1 MRYDEVIAFLGEMGFQRLIFPLLSASTIIPNCFNMSVYFLAGPHEHCRRPDANISS 60  
 DB 1 MRYDEVTAFLGEMGFQRLIFPLLSASTIIPNCFNMSVYFLAGPHEHCRRPDANISS 60  
 QY 61 AMRNNSVPLRLRGREVPNHSRRLATIANFSALGLEGRVDLQGLQESCLDGMERS 120  
 DB 61 AMRNNSIPLRLETKGRQVPOKCRKRLATIANFSSELGRVDLQGLQESCLDGMERYD 120  
 QY 121 QDYVLTSTVTENMLVCEDMKVPPLTTSLEFVGVLGSEVSGQLSDRGKRVLFATMAVQ 180  
 DB 121 KDVEFLSTIVEMDLVCKDKMAPLTTSLEFVGVLGSEVSGQLSDRGKRVLFATMAVQ 180  
 QY 181 TGFSFQIPIISIMEMETVLFVYGMGOISNYVAFLTGELLSKSRILFFATLVCIFFA 240  
 DB 181 TGFSFQIVSVNEMETVLFVYGMGOISNYVAFLTGELLSKSRILFFATLVCIFFA 240  
 QY 241 VGMPLPLFAVFRDMRLMLALTVPGLVPLMFIPEPSRMLISQRFREAEITOKA 300  
 DB 241 FGMVPLPLFAVFRDMRLMLALTVPGLVPLMFIPEPSRMLISQRFREAEITOKA 300  
 QY 301 AKANNATVPAVITD--SVEELNPLKQKAFILDFETRNAIMTMSLLMLMTSGYFA 358  
 DB 301 AKINGIVAPSTIPDPELDLNTKPOLHIIYDLIRINIRIVITMSIILMTTISVGYG 360  
 QY 359 LSLDAPLHDDAVLNFELSLAIEIPAYITAMLLRTPRYITAMLLFNGGGLFETOLY 418  
 DB 361 LSLDTPHLDIYVNFELSLAIEIPAYITAMLLRTPRYITAMLLFNGGGLFETOLY 420  
 QY 419 PVDYFELSLGLVNLGFGITSAFSLMIVFTALPTLVARNMAVGVTSAVSGSIITAPYF 478  
 DB 421 PSELFLSTALVWVGFGITSAFSLMIVFTALPTLVARNMAVGVTSAVSGSIITAPYF 480  
 QY 479 VYLGAYRMLPYVMSLTVLIGITFLFPESGMLTPELEGMKVKKFRSGK--KTR 535  
 DB 481 VYLGAYDRFLPYLMSLTVLITLITLFPESGVPPLPTIDMLRVKIKQMIOIOSQTR 540  
 QY 536 DSMETENPKVL-ITAF 551  
 DB 541 MQMDRESPTVLTAKSTAF 557  
 RESULT 4  
 YLXS CAEEL STANDARD; PRT; 751 AA.  
 ID YLXS CAEEL  
 AC P46501;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, last sequence update)  
 DT 01-NOV-1995 (Rel. 32, last annotation update)  
 DE Hypoetical 84.8 kDa protein F23f12.5 in chromosome III.  
 GN F23f12.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Du Z.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U12965; AAA20607.1; -  
DR Wormpep: F23F12.5; CE01252.  
DR InterPro: IPR002184; Sfb.  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF02175; Sfb; 1.  
DR Pfam: PF00083; sugar\_tr; 1.  
KM Hypothetical protein; Transmembrane.  
FT TRANSMEM 24 44 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT TRANSMEM 102 122 POTENTIAL.  
FT TRANSMEM 137 157 POTENTIAL.  
FT TRANSMEM 200 220 POTENTIAL.  
FT TRANSMEM 235 255 POTENTIAL.  
FT TRANSMEM 320 340 POTENTIAL.  
FT TRANSMEM 348 368 POTENTIAL.  
FT TRANSMEM 379 399 POTENTIAL.  
FT TRANSMEM 410 430 POTENTIAL.  
FT TRANSMEM 432 452 POTENTIAL.  
FT TRANSMEM 515 535 POTENTIAL.  
FT TRANSMEM 547 567 POTENTIAL.  
FT TRANSMEM 583 603 POTENTIAL.  
FT TRANSMEM 614 634 POTENTIAL.  
FT TRANSMEM 678 698 POTENTIAL.  
SQ SEQUENCE 751 AA; 84832 MW; A6C4F3540295ERC CRC64;

Query Match 13.8%; Score 393; DB 1; Length 751;  
Best Local Similarity 27.0%; Pred. No. 2,4e-18;  
Matches 119; Conservative 91; Mismatches 184; Indels 46; Gaps 14;

OY 127 TVYVEMNLVCEDN-WKVPVLTSLFVGVGLGSPVSGQLSDPFGRKNVLFAMAVOTGFSF 185  
DB 303 SMVODEKFCGCTKADAAMVATVTFQIGVLTGATVYGHGDFGRKPVSFGISVGLLFGV 362  
OY 186 LQIFSIWEMFTLVFIYVGMQISNYVAFILGTEILGKSVRIIFSTLGVCTFPFVAVG-- 243  
DB 303 ASGFAPSMVEFAAFRFIVGTSLILVYAYILFEIPEQRFVLR-----SFFNMGYAR 417  
OY 244 MLPLFAFYIRDMRLMLLALTPVGLVPLWFIPESEPRMLISORPREAEDIIOKAKM 303  
DB 418 LVFTLACFICGVMRSAAIATSLILPLIPVLLILPESEPKMENTKKRFADARAARAVAML 477  
OY 304 NNTANPAY-----IDSVBELNPLKQOKAFIL-DIFRRNTAINTIMSLMLLSV 354  
DB 478 --SGIPVYNDEQSIETSEKLEE---KSTRIYTKMDLFTSWTAYRIIVGSLFSTSL 531  
OY 355 GYFALSLDAPMLHGDVAYNCFSLALIEIPAYITAMLLRTLP---RRIY-----IAAVLF 406  
DB 532 SAFSGDLMSGNLAGNFYISQFVSGAVTAFAKIFVPLDITVPSDRRLHYPQIAMLIC 591  
OY 407 WGGVGLFIQLVPV-----DYFLSIGVLMGKFGITSAPSMILVFTAEIYPTLVR 457  
DB 592 Y--CVIWMILMIPESDCSGSQRDAIITINIGVSFEIT--WDACYLVAVEGPFTRIR 647  
OY 458 NNNAVGVTSASVSGSIIPFYVYLGAVYRMPLPYIMGSL--TVLIGITFLFEPESIGMPLP 516  
DB 648 TIGITCSLNTGALLPQANVYISDIYRAPYAVVCSIGTISLISCVLPDTRKVDL- 706  
OY 517 ETLEQOMAKVFRSGKTRD 536  
DB 707 AALDPTLELDYDRKKSMEN 726

RESULT 5

YMP3\_CAEEL  
ID YMP3\_CAEEL STANDARD; PRT; 1222 AA.  
AC 010947;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 139.9 kDa protein B0361.3 in chromosome III.  
GN B0361.3  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
RN NCBI\_TaxID=6239;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Du Z.;  
RL Submitted (FEB-1994) to the EMBL/Genbank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U00031; AAK1864.1; -  
DR Wormpep: B0361.3; CE00752.  
DR Pfam: PF00083; sugar\_tr; 1.  
KM Hypothetical protein; Transmembrane.  
FT TRANSMEM 51 71 POTENTIAL.  
FT TRANSMEM 148 168 POTENTIAL.  
FT TRANSMEM 182 202 POTENTIAL.  
FT TRANSMEM 232 252 POTENTIAL.  
FT TRANSMEM 263 283 POTENTIAL.  
FT TRANSMEM 307 327 POTENTIAL.  
FT TRANSMEM 329 349 POTENTIAL.  
FT TRANSMEM 355 375 POTENTIAL.  
FT TRANSMEM 382 402 POTENTIAL.  
FT TRANSMEM 413 433 POTENTIAL.  
FT TRANSMEM 442 462 POTENTIAL.  
SQ SEQUENCE 1222 AA; 139868 MW; CBA42A80A254FB80 CRC64;

Query Match 11.2%; Score 317.5; DB 1; Length 1222;  
Best Local Similarity 20.3%; Pred. No. 3e-13;  
Matches 114; Conservative 106; Mismatches 219; Indels 123; Gaps 14;

OY 3 DYDEVINFLGEMGPF-----QRLIFLLSIIIPNGFGKSVVFLAGTPEHRCVDP 54  
DB 35 DPKRFVAYATYKQYQFTYVLYVQTLNLFYSSMYI-----MSFYQL--NLEKQE---- 83  
OY 55 AANLSAMRNNSVPLRLDRGEPVHSCSRYRIATIANFSALGLEPGRVDLQOLEQESCL 114  
DB 84 -----YKNETIP-----ISETCQ-----IETSSKAFGLNMEYC- 113  
OY 115 DGMFESQDVL-----STVYTEMNLVCEDNWKVPLTSLFVGVGLGSPVSGQLSQR 166  
DB 114 --GIAENTLVNVTNOKASNTLVDPDLSCHWFQEFQGITITIGAVIAVPMMLADR 170  
OY 167 FGRKNVVFATMAVOTGFSFQIIFSIWEMFTLVFIYVGMQISNYVAFILGTEILGKSV 226  
DB 171 YGRKPIIVTALLAFLANMAASFPNAPILILIRAFIGACSDSLVSAVATGEYSEKA 230  
OY 227 RIIFSTLGVCTFPFVAVGYMLPLFAFYIRDMRLMLLALTPVGLVPLWFIPESEPRMLIS 286  
DB 231 R-AMITVVYVAVMSIGVWVTLVLTMTDWMRYRYFTVSLPGVGFALWYFLPESPHMLIT 289  
OY 287 QRRPREAEDIIOKAAKNNNAVPAVIFDSVEELNPLKQOKAFILDLFRRNTAINTIMSL 346  
DB 290 KNTKTKLYIKTANRM-----VISL 310



RESULT	3
YCEL_BACSU	
ID	YCEL_BACSU
AC	034691;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Hypothetical metabolite transport protein ycel.
GN	YCEL.
OS	<i>Bacillus subtilis</i> .
OC	Bacteria; Firmicutes; <i>Bacillus</i> / <i>Clostridium</i> group;
OC	<i>Bacillus</i> / <i>Staphylococcus</i> group; <i>Bacillus</i> .
OX	NCBI_TaxID=1423;
RN	(1)
RP	SEQUENCE FROM N.A.
RC	STRAIN=168;

```

0Y 116 GWEE-SQDY-YLSTVY- ---DEMNVED- NMWYPLTTSIFVGLSPSGQSLSRFG 168
Db 19 GWLEFDADVGIIISFIIALHVEWMLSPBEKKW- ---IGVSNISGMAGAFLEGLADRIG 74
0Y 169 KKNVLPATMAVOTGSEFLOIRSISHEMFVLYFVYMGQISNVAFFILGEIL- ---G 2233
Db 75 RKKRFEITTLCSIGSISGISAFTLSAFLTLRFVYISGJGGLBPAASTVSDAVVPEKRG 1344
0Y 224 KSVRIIFSTIGVCTEPAVGMYMLPLFAEFI- ---RDMRMLLALTPVGLCVLPMWFIPES 2804
Db 135 RVILTES- ---FWAGWMLAALISYFVIPRSGWQALLLTALTALPAFVLYLRTSLPDS 1888
0Y 281 PRMLISRRPREADELIQKAAKMMNTAVPAVIDPSYELNPLKQOKAFILDLFTRNIAI 3404
Db 189 PKY- ---ESLSAKRSM- ---MENKSVY- ---WAKOYIR 2155
0Y 341 MTIISLLMLMTISVGFALSLDAPN- ---LHGDAYLWCF- ---LSALIEIPAYITAMDLR 3933

```

Db 216 PTVMLSIWFECVFSYGMFLMPLSVMLKGFMSIOSEFVLLMTLAOLPGFSAAMLE 275  
 QY 394 TLPRRTIAAVLFMGCGVLLFLQLVPDYDFLSIGVLMLKFGITSAFSLMVFATLPR 453  
 Db 276 KAGRKWLVVYLGITAGSAHYFPGTADISLLTLTG-VLSEFFNU-GAMGVLYAYTEQYR 333  
 QY 454 TLVRNANGVSTASRVSGSIAPFYVYLGAVNRMPLPYVMSLTLVIGITFL 505  
 Db 334 TAIRATGSGTFAFGRIIGIFGPLVGTLAARH-----SFSVIFSEICI 378  
 RESULT 9  
 YAU\_ECOLI STANDARD: PRT: 443 AA.  
 ID YAU\_ECOLI  
 AC P31679; P31578; P75628;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1999 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical metabolite transport protein yau.  
 GN YAU OR B0045.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=92334977; PubMed=1630901;  
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,  
 RA Isono K., Mizobuchi K., Nakata A.;  
 RT "Systematic sequencing of the Escherichia coli genome: analysis of  
 RT the 0-2.4 min region."  
 RL Nucleic Acids Res. 20:3305-3308(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12."  
 RL Science 277:1453-1474(1997).  
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (potential).  
 CC -1 SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -1 CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS  
 CC IN POSITION 142 AND 232 THAT PRODUCE TWO SEPARATE ORFS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D10483; BAA01321.1; ALT\_FRAME.  
 DR EMBL: AE000114; AAC73156.1; -.  
 DR PIR: S40566; S40566.  
 DR Ecogene: EG11566; yau.  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF00083; sugar\_tr.1.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
 KW Hypothetical protein; Transport; Transmembrane; Inner membrane;  
 KW Complete proteome.  
 FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 19 39 1 (POTENTIAL).  
 FT DOMAIN 40 53 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 54 74 2 (POTENTIAL).  
 FT DOMAIN 75 84 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 85 105 3 (POTENTIAL).

FT DOMAIN 106 113 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 114 134 4 (POTENTIAL).  
 FT DOMAIN 135 145 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 146 166 5 (POTENTIAL).  
 FT DOMAIN 167 173 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 174 194 6 (POTENTIAL).  
 FT DOMAIN 195 241 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 242 262 7 (POTENTIAL).  
 FT TRANSMEM 263 282 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 283 303 8 (POTENTIAL).  
 FT TRANSMEM 304 309 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 310 329 9 (POTENTIAL).  
 FT TRANSMEM 330 334 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 335 357 10 (POTENTIAL).  
 FT TRANSMEM 358 373 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 374 394 11 (POTENTIAL).  
 FT TRANSMEM 395 401 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 402 422 12 (POTENTIAL).  
 FT DOMAIN 423 443 CYTOPLASMIC (POTENTIAL).  
 FT CONFLICT 423 443 F->L (IN REF. 1).  
 FT CONFLICT 12 16 FSSH->CLIH (IN REF. 1).  
 FT CONFLICT 30 30 D->Y (IN REF. 1).  
 FT CONFLICT 115 115 L->F (IN REF. 1).  
 FT CONFLICT 132 132 T->N (IN REF. 1).  
 FT CONFLICT 140 240 R->C (IN REF. 1).  
 FT CONFLICT 304 304 N->S (IN REF. 1).  
 FT CONFLICT 395 395 N->D (IN REF. 1).  
 SQ SEQUENCE 443 AA; 48666 MW; AEF40BD8109DE05 CRC64;  
 Query Match 9.6%; Score 274.5; DB 1; Length 443;  
 Best Local Similarity 25.8%; Pred. No. 6.2e-11;  
 Matches 101; Conservative 76; Mismatches 167; Indels 47; Gaps 12;  
 QY 152 GVLGSEFVGQSDRGRKRVLFATMAVGQFSLQJFISMEWETLYFYVGMQGISNV 211  
 Db 65 GLFVGTSLGYSIDVGRKKMFLIDIAIVISVATMEVSSPELLMRLVIGIVGADY 124  
 QY 212 VVAFLGTETLKSVAIIFSTLGVCTFPVGYMLPLFAFYIRD---WRMLLALTVPG 267  
 Db 125 PIATMITEFSSTROR-AFISIFIAAMVYGATCADLVGYWLVDEGGMMLGSAIIC 183  
 QY 268 YLCVPLMFIIPSPRMLISQRFREAEEDIIKAKMNTAVPAVPIPSVELNPLQOQA 327  
 Db 184 LTLIGREFLPSPKMLKGRKCEBEMTK-----LREPVAFDEDPQOT 231  
 QY 328 FILDRTNINAMTMSLLMLTSVGFALSIDAPNLG-----DAYLNCFLSL 379  
 Db 232 RFRDLFNRHFPFLVFA-AIWTQVIMPAIYFPGQIVGLGLGKMAALGNVVISL 290  
 QY 380 IETPAITVMLLRTLPRTIYAANVLFMGGVLLFLQLVPDYDFLSIGVLMLKFGITS 439  
 Db 291 EFMIGCIIPMLMLNTRGRPRPLIGSFAMMTLALVGLIP---DMGIMLVNA-FAVYA 345  
 QY 440 AFS-----MLVYFTELYPTIVRNNAVGVSTASRVSGSIAPFYVYLGAVNRMPL 489  
 Db 346 FFGSGPGNQLWLYPELPTDIDIRASAVGVIMSLRIGIVSTALPRLINNVGISVTML- 404  
 QY 490 YIMGSLTVLIG--FTLFPFESLGMTLPER 518  
 Db 405 ---WGAGISLFGLLISVAFAPETRGMSLAQT 432  
 RESULT 10  
 CSBC\_BACSU STANDARD: PRT: 461 AA.  
 ID CSBC\_BACSU  
 AC P46333; O32288;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Probable metabolite transport protein csbc.  
 GN CSBC OR SS928R.  
 OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_Taxid=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / BGSC1A1;  
 RX MEDLINE=96093926; PubMed=7584049;  
 RA Yoshida K.-I., Seki S., Fujimura M., Miwa Y., Fujita Y.,  
 RT "Cloning and sequencing of a 36-kb region of the Bacillus subtilis  
 RL genome between the gnt and tol operons.";  
 RN DNA Res. 2:61-69(1995).  
 [2]  
 RP REVISIONS.  
 RA Fujita Y., Shibayama T., Ishio I., Aoyama D., Yoshida K.-I.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP CHARACTERIZATION.  
 RX MEDLINE=99303315; PubMed=10376822;  
 RA Akbar S., Lee S.Y., Boylan S.A., Price C.W.;  
 RT "Two genes from Bacillus subtilis under the sole control of the  
 RL general stress transcription factor sigmaB.";  
 CC Microbiology 145:1069-1078(1999).  
 CC -1- FUNCTION: COULD SERVE EITHER A NUTRITIONAL OR AN OSMOTIC  
 CC PROTECTION FUNCTION.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB005554; BAA21604.1; -;  
 DR EMBL: 299124; CAB16017.1; -;  
 DR Subtilist; BG1360; csbc.  
 DR InterPro: IPR003663; Sugar\_transporter.  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF00063; sugar\_tr; 1.  
 DR PRINTS: PR00171; SUGRTNSPORT.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 2.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
 DR Transport; Transmembrane; Complete proteome.  
 FT DOMAIN 1 14  
 FT TRANSMEM 15 14  
 FT TRANSMEM 15 35  
 FT DOMAIN 36 38  
 FT TRANSMEM 39 59  
 FT DOMAIN 60 76  
 FT TRANSMEM 77 97  
 FT DOMAIN 98 104  
 FT TRANSMEM 105 125  
 FT DOMAIN 126 139  
 FT TRANSMEM 140 160  
 FT DOMAIN 161 163  
 FT TRANSMEM 164 184  
 FT DOMAIN 185 241  
 FT TRANSMEM 242 262  
 FT DOMAIN 263 280  
 FT TRANSMEM 281 301  
 FT DOMAIN 302 308  
 FT TRANSMEM 309 329  
 FT DOMAIN 330 341  
 FT TRANSMEM 342 362  
 FT DOMAIN 363 378  
 FT TRANSMEM 379 399  
 FT DOMAIN 400 402

FT TRANSMEM 403 423  
 FT DOMAIN 424 461  
 FT CONFLICT 400 401  
 SO SEQUENCE 461 AA; 50235 MW; B6587B5DC272EF1 CRG64;  
 Query Match 9.5%; Score 270; DB 1; Length 461;  
 Best Local Similarity 26.3%; Pred. No. 1.3e-10;  
 Matches 118; Conservative 79; Mismatches 180; Indels 72; Gaps 19;  
 QY 139 NMKVPFT-----SLFVGVLGSGVSGQSDRGKRNVEFATMAVGTGSPFIQFSI 191  
 DB 36 NNDIPFTLTLEGLVSMILGAIFGSALSTCSDRMKRRVVFVLSITITGALCAFSQ 95  
 QY 192 SWMEFTVLFVYGQISNVVAFFILGTLLGKSVRIIFSTL-----GYCTFPAVGVM 244  
 DB 96 TIGMLISRVILGLAVGSTALVPVYLSEMAPKIRGTGMNMLIVGILLAVIVNL 155  
 QY 245 LLPLFAFIDWRMLLALVPGVLCVPLMWFPESPRGLISGRREDEDIQAARKN 304  
 DB 156 FTPEEA-----WRMWGLAAPVALLIGIAFPEPRMLVKRGSSEARRI-----MN 204  
 QY 305 NTAAPAVIFPSVEELNPKQ---OKAFILDFERTNIAIMTMSL-LTMLTSVGYPAL 359  
 DB 205 ITHDPK---DIEMELAMKQGEAKKETTGLVAKAWIRPMLLIGVGLAIFQAVGINTV 261  
 QY 360 SLDPNHLGDAVLCNCFSL-----IEIPAVITWMLRLPRTYITAAVLFMGGYL 412  
 DB 262 IYVAPFTFRAGLGTSSALGTMTGIGLIVMCTIMAILIDRGRK---KLIM-GSYG 316  
 QY 413 LFIOLVVDVYFSLIGL-----VMLKRGFI---TSAFSLVFTALPYTLVRNMA 460  
 DB 317 ITLSLALSGVLLTLGLSASTAMTVELGVIVFQATMGVPVWMLPELPSKARGAA 376  
 QY 461 VGVP-----STRSRVGSIIAFYVY-LGAVNRMPLIVM-GSLTVLIGITLFF--PESLG 512  
 DB 377 TGFTTLVLSANLIVSLVFPILMRPMG-----TAWFMVFSVLCISFPFAFVMPETKG 431  
 QY 513 MLEPETLEOMQKVKWFRSGKTRDSMTE 541  
 DB 432 KSL-EEIEASLK-KRFKKKSTQNOVINE 458  
 RESULT 11  
 YGCS\_ECOLI STANDARD; PRT; 445 AA.  
 ID YGCS\_ECOLI  
 AC 046909;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical metabolite transport protein ygcS.  
 GN YGCS OR B2771.  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 OX NCBI\_Taxid=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 CC (potential).  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its

ID	GTRL_CHKCK	STANDARD:	PRT:	490 AA.
RESULT 12				
AC	P46896;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Solute carrier family 2, facilitated glucose transporter, member 1			
DE	(Glucose transporter type 1) (GTL).			
GN	SIC2A1 OR GLUT1.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OX	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96157892; PubMed=8589457.			
RA	Wagstaff P., Kang H.Y., Mylotte D., Robbins P.J., White M.K.;			
RT	"Characterization of the avian GLUT1 glucose transporter:			
RT	differential regulation of GLUT1 and GLUT3 in chicken embryo			
RL	fibroblasts.";			
RL	Mol. Biol. Cell 6:1575-1589(1995).			
CC	-1 FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. THIS ISOFORM MAY BE			
CC	RESPONSIBLE FOR CONSTITUTIVE OR BASAL GLUCOSE UPTAKE. HAS A VERY			
CC	BROAD SUBSTRATE SPECIFICITY; CAN TRANSPORT A WIDE RANGE OF ALDOSES			
CC	INCLUDING BOTH PENTOSE AND HEXOSE (BY SIMILARITY).			
CC	-1 SUBCELLULAR LOCATION: Integral membrane protein. Localizes			
CC	primarily at the cell surface (by similarity).			
CC	-1 SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE			
CC	TRANSPORTER SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; L07300; AAB02037.1; "			
DR	InterPro; IPR003663; Sugar transporter.			
DR	InterPro; IPR003662; sub_transporter.			
DR	Pfam; PF00083; sugar_tr_1			
DR	PRINTS; PR00171; SUGKTRNSPORT			
DR	PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.			
DR	PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.			
KW	Transmembrane; Sugar transport; Transport; Glycoprotein;			
KW	Multigene family.			
FT	DOMAIN 1	11	CYTOPLASMIC (POTENTIAL)	
FT	TRANSSEM 12	32	1 (POTENTIAL)	
FT	DOMAIN 33	65	EXTRACELLULAR (POTENTIAL)	
FT	TRANSSEM 66	86	2 (POTENTIAL)	
FT	DOMAIN 87	94	CYTOPLASMIC (POTENTIAL)	
FT	TRANSSEM 95	115	3 (POTENTIAL)	
FT	DOMAIN 116	125	EXTRACELLULAR (POTENTIAL)	
FT	TRANSSEM 126	146	4 (POTENTIAL)	
FT	DOMAIN 147	154	CYTOPLASMIC (POTENTIAL)	
FT	TRANSSEM 155	175	5 (POTENTIAL)	
FT	DOMAIN 176	184	EXTRACELLULAR (POTENTIAL)	
FT	TRANSSEM 185	205	6 (POTENTIAL)	
FT	DOMAIN 206	270	CYTOPLASMIC (POTENTIAL)	
FT	TRANSSEM 271	291	7 (POTENTIAL)	
FT	DOMAIN 292	306	EXTRACELLULAR (POTENTIAL)	
FT	TRANSSEM 307	327	8 (POTENTIAL)	
FT	DOMAIN 328	336	CYTOPLASMIC (POTENTIAL)	
FT	TRANSSEM 337	357	9 (POTENTIAL)	
FT	DOMAIN 358	369	EXTRACELLULAR (POTENTIAL)	
FT	TRANSSEM 370	390	10 (POTENTIAL)	
FT	DOMAIN 391	400	CYTOPLASMIC (POTENTIAL)	
FT	TRANSSEM 401	421	11 (POTENTIAL)	





```

CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; AE000272; AAC74845.1; -
DR      EMBL; D90821; BA013573.1; ALU_INIT.
DR      EcoGene; EG13487; ydjK.
DR      InterPro; IPR003662; sub_transporter.
DR      Pfam; PF00083; sugar_tr; 1.
DR      PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
DR      PROSITE; PS00217; SUGAR_TRANSPORT_1; FALSE_NEG.
KW      Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW      Complete proteome.
FT      DOMAIN 1 25 CYTOPLASMIC (POTENTIAL)
FT      TRANSMM 26 46 1 (POTENTIAL)
FT      DOMAIN 47 60 PERIPLASMIC (POTENTIAL)
FT      TRANSMM 61 81 2 (POTENTIAL)
FT      DOMAIN 82 90 CYTOPLASMIC (POTENTIAL)
FT      TRANSMM 91 111 3 (POTENTIAL)
FT      DOMAIN 112 112 PERIPLASMIC (POTENTIAL)
FT      TRANSMM 113 133 4 (POTENTIAL)
FT      DOMAIN 134 153 CYTOPLASMIC (POTENTIAL)
FT      TRANSMM 154 174 5 (POTENTIAL)
FT      DOMAIN 175 181 PERIPLASMIC (POTENTIAL)
FT      TRANSMM 182 202 6 (POTENTIAL)
FT      DOMAIN 203 271 CYTOPLASMIC (POTENTIAL)
FT      TRANSMM 272 292 7 (POTENTIAL)
FT      DOMAIN 293 301 PERIPLASMIC (POTENTIAL)
FT      TRANSMM 302 322 8 (POTENTIAL)
FT      DOMAIN 323 329 CYTOPLASMIC (POTENTIAL)
FT      TRANSMM 330 350 9 (POTENTIAL)
FT      DOMAIN 351 351 PERIPLASMIC (POTENTIAL)
FT      TRANSMM 352 372 10 (POTENTIAL)
FT      DOMAIN 373 399 CYTOPLASMIC (POTENTIAL)
FT      TRANSMM 400 420 11 (POTENTIAL)
FT      TRANSMM 421 441 12 (POTENTIAL)
FT      DOMAIN 442 459 CYTOPLASMIC (POTENTIAL)
SQ      SEQUENCE 459 AA; 49602 MW; BC8AB53ECB8BD077 CMC64;

Query Match 8.6%; Score 245; DB 1; Length 459;
Best Local Similarity 23.4%; Pred. No. 5,1e-09;
Matches 97; Conservative 81; Mismatches 177; Indels 60; Gaps 17;

QY 134 LVCEN-----WKVLTITSLF-----FVGLDSFVSQGLSDRGRKNVLA 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 LVCNMAVGLILQALKGWTDNSTVFAITTAGFALVGGITGDTGRNRAFL 93
QY 176 TMAVOTGRSFLOIFISIMEMFTVLVIYGMQISNYVAFLIGTEIL-GK-----SVRII 229
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 94 YEALHISMVVGAGSPNNDLLACRFVAGLALYLVLFAGFTIYMGRRKGTSSKVS 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 FS---TLGVTCTPAVGVMLPLFAFFIRDMRMALLATVPGVLCPL-WMTIPESRWLI 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 FIGMWSYDLCSLIMAG--LTPPLIS-AEWMNVQQLILPAILSLIATAIAMARFPESPWLE 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 SORREAREDEIIOK---AAKMNNTAVAVAFEDSVEELNPLKQOKAFILDEFRFNTI--- 338
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 211 SKRGIOEEKWKRSTIEGCVIRQTKGKPLRPVVIADGKAPQAVPYSAITLVGLKRVILGS 270
QY 339 AAKMTMSLLMLMLTSVGFALSLDAPNHDGAYINCFALIEIP-AVITAMLLRTLP 397
   : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 271 CVLLAMNVQYTLTNMLPTIFMTOGINIKDSIVLNTM--SMFGAPFCFIAMLVMDKIPR 328
   : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 398 RYITAAVLFMGGVLLPTQLVDPVDYIELS--IGLVMGKFGSITAFSKLY-----VFTA 449
   : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 329 KTM-----GVGLLILAVLGVIYSLOTSMILLITLIGFELIT--FVMYCVASAVYVP 379

```

OY 450 ELPPLVNNNAVGVSTASRVGSIAPY--PYIGAYNRMLPYIWMGSLTAVLIGI 502  
 Db 380 EIMPTFALRGSGLANAVGRISGIAAPYAAVALLSYGVGVGTFILLDAVSLITVAI 434

RESULT 15  
 KHT2\_KLUUA STANDARD: PRT: 566 AA.  
 AC P53387;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hexose transporter 2.  
 GN KHT2.  
 OS Kluyveromyces fragilis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-JAB;  
 RA MEDLINE=98028406; PubMed=9363776;  
 RX Weirich J., Goffin P., Kuger P., Ferrero I., Breunig K.D.;  
 RT "Influence of mutations in hexose-transporter genes on glucose  
 repression in Kluyveromyces fragilis".  
 RL Eur. J. Biochem. 249:248-257(1997).  
 CC -1- FUNCTION: PROBABLE GLUCOSE TRANSPORTER.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: 247080; CAA87389.1; -  
 DR InterPro: IPR003663; Sugar\_transporter.  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF00083; sugar\_tr.1.  
 DR PRINTS: PR00171; SUGRTRANSPORT.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
 KW Repeat; Transmembrane; Sugar transport; Glycoprotein.  
 KM DOMAIN 22 61  
 FT TRANSMEM 62 82 1 (POTENTIAL).  
 FT DOMAIN 83 112 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 113 133 2 (POTENTIAL).  
 FT DOMAIN 134 140 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 141 161 3 (POTENTIAL).  
 FT DOMAIN 162 166 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 167 187 4 (POTENTIAL).  
 FT DOMAIN 188 198 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 199 219 5 (POTENTIAL).  
 FT DOMAIN 220 233 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 234 254 6 (POTENTIAL).  
 FT DOMAIN 255 333 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 334 353 7 (POTENTIAL).  
 FT DOMAIN 354 357 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 358 378 8 (POTENTIAL).  
 FT DOMAIN 379 385 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 386 406 9 (POTENTIAL).  
 FT DOMAIN 407 428 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 429 449 10 (POTENTIAL).  
 FT DOMAIN 450 465 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 466 486 11 (POTENTIAL).  
 FT DOMAIN 487 492 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 493 513 12 (POTENTIAL).  
 FT DOMAIN 514 566 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 88 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SO SEQUENCE 566 AA; 62727 MW; 9784173EC3735444 CRC64;

Query Match 8.6%; Score 245; DB 1; Length 566;  
 Best Local Similarity 22.2%; Pred. No. 6.4e-09;  
 Matches 106; Conservative 78; Mismatches 196; Indels 98; Gaps 17;

OY 139 NMKVPLTTSLEFFYGVLLGSFVSQSLDRGKKNVLPATNAVGTGSEFLOIFS1-SWEMET 197  
 Db 109 NVRGTGLVIFENIGCAIGILLSKIDMGRRIGIMIVLIVVGGIIQIASIDKMYQYF 168  
 OY 198 VLEFVVGQISNIVVAFLIGTELKSVIITSTIGCVTFPFAVGMILPFAVIR--- 254  
 Db 169 IGRISGLGVGGISVLSPLMISTEPARKIR-----GLVVF---YQMTTEGIFLGCT 219  
 OY 255 -----DPRMLLALTVPGLCVPLWMFIPESPRLISQRRFEADITOKAAM 303  
 Db 220 NYGKITYSNSVOMRVPLGLCFANAIFMITGMLVPEPSRFLVCKDRIDEAKRSIAKSNV 279  
 OY 304 N--NTAVPA---VIFDSVEELNPLKQKAPILDLFTR-NAIMTMSLL--WMLTSV 354  
 Db 280 SYEDPAVQAEVDLICAGVEA--ERLAGSASIKELFSTKVFQRLIMGMLIOSFOOLTGN 337  
 OY 355 GYF-----ALSLAPNLHGDAVINCFLSLALIEIPATITMILLRLPRRYIIAAVY 405  
 Db 338 NYFFYGTITFNSVGM-----DSFETSIVLGVNFASIFVAIYVDKGR--KCL 387  
 OY 406 FMGGGVLLFQLVVDYFFLSIGLVMGKFG-----ITS 439  
 Db 388 LMGAAMATACNV-----FASVGTRLMPDGANHPETASGAGNCMTVFACFIICGATS 442  
 OY 440 AFSLMYFTAEPLVTLVNNNAVGVSTASRVGSIAPYV--YIGAYNRMLPYIWMGSLT 497  
 Db 443 WAPIAYVVAASYPPLRYKAKMAIATASNIMWGLNFFFPITSAIHFFYGVFMCLV 502  
 OY 498 VLIIGTIFPESIGMTLPETLEOMK-----VKMFRSGKKR---DSMETEENP 544  
 Db 503 AMFYVFFFEVETGLTLEVOEMMERGVLPWKSSWVPSRRNAGYDVADLOHDEXP 560

Search completed: July 17, 2002, 02:15:01  
 Job time: 662 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 02:03:14 ; Search time 102.99 Seconds

(without alignments)  
925.529 Million cell updates/sec

Title: US-09-521-195b-1

Perfect score: 2845

Sequence: 1 MRDYEVIATFGEWGPFORL.....KTRDSMETENPKVLITAF 551

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2845	100.0	551	4	014546
2	2833	99.6	551	4	09H015
3	2484	87.3	553	11	09R141
4	2470	86.8	553	11	09Z306
5	2211	77.7	557	4	096EH6
6	2034.5	71.5	564	11	09WTN6
7	857	30.1	548	5	09YCA2
8	852	29.9	548	5	001384
9	819	28.8	567	5	09YCA3
10	815	28.6	567	5	095R48
11	794.5	27.9	577	4	096R00
12	781.5	27.5	563	5	095TW6
13	773.5	27.2	561	5	09Y6L6
14	754.5	26.5	568	5	09Y539
15	754.5	26.5	576	5	002270
16	709	24.9	540	11	091W02

17	707	24.9	554	6	077504	077504 oryctolagus
18	703	24.7	555	4	015244	015244 homo sapien
19	700	24.6	554	6	002713	002713 sus scrofa
20	699.5	24.6	593	11	P70485	P70485 rattus norv
21	698.5	24.6	604	5	0961J5	0961J5 drosophila
22	696.5	24.5	553	11	070577	070577 mus musculu
23	695.5	24.4	555	11	09R0M2	09R0M2 rattus norv
24	695	24.4	674	5	09YIK2	09YIK2 drosophila
25	692	24.3	556	11	008966	008966 mus musculu
26	691.5	24.3	593	11	P97558	P97558 rattus norv
27	688.5	24.2	554	4	09N0D4	09N0D4 homo sapien
28	688.5	24.2	554	4	015395	015395 homo sapien
29	687.5	24.2	556	11	063089	063089 rattus norv
30	687	24.1	556	11	09R1O4	09R1O4 mus musculu
31	684.5	24.1	554	4	015245	015245 homo sapien
32	673	23.7	535	11	063314	063314 rattus norv
33	673	23.7	555	5	0961R9	0961R9 drosophila
34	672	23.6	557	5	09YEX8	09YEX8 drosophila
35	671	23.6	556	4	075751	075751 homo sapien
36	662.5	23.3	557	4	096KH7	096KH7 homo sapien
37	653.5	23.0	456	4	014567	014567 homo sapien
38	653.5	23.0	548	4	09Y694	09Y694 homo sapien
39	652	22.9	539	4	09H2W5	09H2W5 homo sapien
40	651	22.9	542	4	096TC1	096TC1 homo sapien
41	650	22.8	551	11	09WTW5	09WTW5 mus musculu
42	648.5	22.8	575	5	09BKR9	09BKR9 caenorhabdi
43	648	22.8	551	4	09Y226	09Y226 homo sapien
44	648	22.8	551	11	099JF0	099JF0 mus musculu
45	640	22.5	551	11	088446	088446 rattus norv

## ALIGNMENTS

RESULT 1  
ID 014546 PRELIMINARY; PRT; 551 AA.  
AC 014546;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JUN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE OCTN1.  
GN OCTN1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE=98086199; PubMed=9426230;  
RA Tamai I., Yabuuchi H., Nezu J., Sai Y., Oku A., Shimane M., Tsuji A.;  
RT cation transporter, OCTN1.";  
RL FEBS Lett. 419:107-111(1997).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
DR EMBL: AB007448; BAA23356.1; -;  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF00083; Sugar\_tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
KW Transmembrane.  
SO SEQUENCE 551 AA; 62177 MW; F5903421C789F60A CRC64;

Query Match 100.0%; Score 2845; DB 4; Length 551;

Best Local Similarity 100.0%; Pred. No. 7.5e-200; Mismatches 0; Indels 0; Gaps 0;

Matches 551; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 MRDYEVIATFGEWGPFORLIFFLISAIIPNGFNGSVYFLAGTPERRRVPDAANLSS 60  
Db 1 MRDYEVIATFGEWGPFORLIFFLISAIIPNGFNGSVYFLAGTPERRRVPDAANLSS 60

Query Match	99.6%	Score 2833	DB 4	Length 551
Best Local Similarity	99.6%	Pred. No. 5.7e-199		
Matches 549	Conservative 0	Mismatches 2	Indels 0	Gaps 0

QY 1 MRDIDEVIAFLGMEGFPORLIFELLASIIENGNGMSVFLACTPEHRCRVDAANLES 60  
 |||||  
 Db 1 MRDYDEVIAFLGMEGFPORLIFELLASIIENGNGMSVFLACTPEHRCRVDAANLES 60  
 |||||  
 QY 61 AMRNNVSPLRLRGCRVPHSCSRVPLATIANPSALGLEPNDVDLGLDEESLDGMS 120  
 |||||

Db	61	AMRNNSPLPLRROGREVPHSCSRRLATIANFSAALGIEPGRDVDLGOLEDESCLDGMEFS	120
Qy	121	QDYLSTVTVEMVLVCEDBNKKVPLTTSLFEVGVLLGSFVSQGLSDRRGRKNVLFATMAVQ	180
Db	121	QDYLSTVTVEMVLVCEDBNKKVPLTTSLFEVGVLLGSFVSQGLSDRRGRKNVLFATMAVQ	180
Qy	181	TGTSFLOIESISNEMFVLVIVGMGQISNRYVAFILGTTELKSNVILFISTLGVCTFEFA	240
Db	181	TGTSFLOIESISNEMFVLVIVGMGQISNRYVAFILGTTELKSNVILFISTLGVCTFEFA	240
Qy	241	VGVMLPLFEAFYFRDMRMMLLATVPGVLCVPLWMFIPESPRMLISQRRPREADILOKA	300
Db	241	VGVMLPLFEAFYFRDMRMMLLATVPGVLCVPLWMFIPESPRMLISQRRPREADILOKA	300
Qy	301	AKNNNTAVPAVIVFDSVEELNPLKOKAFIULDFRTRNIAITMTLSLLMLTVSGYFALS	360
Db	301	AKNNNTAVPAVIVFDSVEELNPLKOKAFIULDFRTRNIAITMTLSLLMLTVSGYFALS	360
Qy	361	LDAPNLHGDAVLYLNCFLSALIEIPAYITAMLLTLPRLRYIAAVLFWGGVLLFIOLVPV	420
Db	361	LDAPNLHGDAVLYLNCFLSALIEIPAYITAMLLTLPRLRYIAAVLFWGGVLLFIOLVPV	420
Qy	421	DYFSLISGLVMGKFGITSFASMLYFETALPYTLVRNNAVGVTASRGSIIAYFVY	480
Db	421	DYFSLISGLVMGKFGITSFASMLYFETALPYTLVRNNAVGVTASRGSIIAYFVY	480
Qy	481	LGAVNMLPYIWNIGSTVLIGITLTFPPESLGMTLEPTLEOMQKVMFNSGKTRDSMET	540
Db	481	LGAVNMLPYIWNIGSTVLIGITLTFPPESLGMTLEPTLEOMQKVMFNSGKTRDSMET	540
Qy	541	EENPKVLITAF 551	
Db	541	EENPKVLITAF 551	
RESULT 3			
Q9RI41		PRELIMINARY;	PRT: 553 AA.
AC Q9RI41:			
DT 01-MAY-2000 (TREMBLrel. 13, Created)			
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE ORGANIC CATION TRANSPORTER OCTN1.			
GN OCTN1.			
OS Rattus norvegicus (Rat).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX NCBI_Taxid=10116;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=20286310; PubMed=10825452;			
RA Wu X., George R.L., Huang W., Wang H., Conway S.J., Leibach F.H.,			
RA Ganapathy V.;			
RT "Structural and functional characteristics and tissue distribution			
RT pattern of OCTN1, an organic cation transporter, cloned from			
RT Placenta.";			
RL Biochim. Biophys. Acta 1466:315-327(2000).			
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).			
CC -I- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.			
DR EMBL: AF169831: AAD46922.1; -			
DR InterPro: IPR003662; sub-transporter.			
DR Pfam: PF00083; sugar_tr.1.			
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.			
KW Transmembrane.			
SO SEQUENCE 553 AA; 62362 MW; E26C8155768A14AD CRC64;			

Query Match	87.3%	Score 2484	DB 11	Length 553
Best Local Similarity	85.2%	Pred. No. 1.7e-173		
Matches 471: Conservative	36	Mismatches 44	Indels 2	Gaps 1
QY	1	MDYDEVATLFGEMGCFORLLIFELLISASIIIPNGFGMSVFLAGTPEHRCRYDPAANTSS	60	



Query Match 77.78; Score 2211; DB 4; Length 557;  
Best Local Similarity 75.88; Pred. No. 1.5e-153;  
Matches 422; Conservative 57; Mismatches 72; Indels 6; Gaps 3;

```
QY 1 MNDYEVIAFLGEMGPFQRLIFFLSASITPNGFNMSVYFAGTPEHRCVPAANLSS 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MNDYEVIAFLGEMGPFQRLIFFLSASITPNGFTGSSVFLIAPENHRCVPAANLSS 60
QY 61 AARNNSVPLRLRDGEVPHSCRYRLATIANFSALGLEGRVDLQGLEQESCLDGMFFS 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 AARNNSVPLRLRDGEVPHSCRYRLATIANFSALGLEGRVDLQGLEQESCLDGMFFS 120
QY 121 ODVYSTVVTETMNLVCEDMKAPLITSLFVGVLLGSFISGOLSDRFGKKNVLFATMAVQ 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 ODVYSTVVTETMNLVCEDMKAPLITSLFVGVLLGSFISGOLSDRFGKKNVLFATMAVQ 180
QY 181 TGFSEFLQFISMEVETVLFYIVGMGOISNYVAFLIGTEILGKSVRIIFSTLGVCTFEFA 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 TGFSEFLQFISMEVETVLFYIVGMGOISNYVAFLIGTEILGKSVRIIFSTLGVCTFEFA 240
QY 241 VGYMLLPFAVFIKDMRLTLATLVPGVLCVPLMFIPESPRMLISQRRFREADIIOKA 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 VGYMLLPFAVFIKDMRLTLATLVPGVLCVPLMFIPESPRMLISQRRFREADIIOKA 300
QY 301 AKANNVAVPAVIFD--SVEELNPLKOOKAFILDFPRRNATIMTSLMLMLTVSGVEFA 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 AKANNVAVPAVIFD--SVEELNPLKOOKAFILDFPRRNATIMTSLMLMLTVSGVEFA 358
QY 359 LSLDAPNHLGDAYLNCFLSALIEIPAYITAMLLRTPRRYIIAAYLVGGGVLLFIOLV 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 361 LSLDAPNHLGDAYLNCFLSALIEIPAYITAMLLRTPRRYIIAAYLVGGGVLLFIOLV 418
QY 419 PDVYFELTFLGVLGKFGITSAFSLVYFAELPTLVNRMAGVSTASRVSITIAFYE 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 421 PDVYFELTFLGVLGKFGITSAFSLVYFAELPTLVNRMAGVSTASRVSITIAFYE 478
QY 479 VYLGAVNMLPYIYMGSLVYIGITFLFPPEISGMLPETLEOMOKVKKFNSGK--KTR 535
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 481 VYLGAVNMLPYIYMGSLVYIGITFLFPPEISGMLPETLEOMOKVKKFNSGK--KTR 535
QY 536 DSEMETEENPKVLTAF 551
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 541 MKKQGERPTLKSTAF 557
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 6  
Q9WTN6 PRELIMINARY; PRT; 564 AA.

AC Q9WTN6; 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE OCTN3.  
GN SLC22A9 OR OCTN3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA MEDLINE=20568258; PubMed=11010964;  
RA Tamai I., Ohashi R., Nezu J., Sai Y., Kobayashi D., Oku A.,  
RT "Molecular and Functional Characterization of Organic Cation/Carnitine  
RT Transporter Family in Mice."  
RL J. Biol. Chem. 275:40064-40072(2000).  
CC -1- SUCCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
DR EMBL: AB018436; BAA78433.1; -.  
DR MGD: MGI:1929481; SLC22A9.  
DR InterPro: IPR003662; sub.Transporter.  
DR Pfam: PF00083; sugar\_tr; 1.

DR PROSITE, PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 564 AA; 63320 MW; C37FDA6395DAD01 CRC64;

Query Match 71.5%; Score 2034.5; DB 11; Length 564;  
Best Local Similarity 69.08; Pred. No. 1.2e-140;  
Matches 380; Conservative 75; Mismatches 85; Indels 11; Gaps 3;

```
QY 1 MNDYEVIAFLGEMGPFQRLIFFLSASITPNGFNMSVYFAGTPEHRCVPAANLSS 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MNDYEVIAFLGEMGPFQRLIFFLSASITPNGFTGSSVFLIAPENHRCVPAANLSS 60
QY 61 AARNNSVPLRLRDGEVPHSCRYRLATIANFSALGLEGRVDLQGLEQESCLDGMFFS 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 AARNNSVPLRLRDGEVPHSCRYRLATIANFSALGLEGRVDLQGLEQESCLDGMFFS 120
QY 121 ODVYSTVVTETMNLVCEDMKAPLITSLFVGVLLGSFISGOLSDRFGKKNVLFATMAVQ 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 ODVYSTVVTETMNLVCEDMKAPLITSLFVGVLLGSFISGOLSDRFGKKNVLFATMAVQ 180
QY 181 TGFSEFLQFISMEVETVLFYIVGMGOISNYVAFLIGTEILGKSVRIIFSTLGVCTFEFA 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 TGFSEFLQFISMEVETVLFYIVGMGOISNYVAFLIGTEILGKSVRIIFSTLGVCTFEFA 240
QY 241 VGYMLLPFAVFIKDMRLTLATLVPGVLCVPLMFIPESPRMLISQRRFREADIIOKA 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 VGYMLLPFAVFIKDMRLTLATLVPGVLCVPLMFIPESPRMLISQRRFREADIIOKA 300
QY 301 AKANNVAVPAVIFDSEVLENDL-----KOOKAFILDFPRRNATIMTSLMLMLTVS 354
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 AKANNVAVPAVIFDSEVLENDL-----KOOKAFILDFPRRNATIMTSLMLMLTVS 354
QY 355 GYFALSLDAPNHLGDAYLNCFLSALIEIPAYITAMLLRTPRRYIIAAYLVGGGVLLFI 414
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 360 GYFALSLDAPNHLGDAYLNCFLSALIEIPAYITAMLLRTPRRYIIAAYLVGGGVLLFI 414
QY 415 IOLVVDVYFELTFLGVLGKFGITSAFSLVYFAELPTLVNRMAGVSTASRVSITIA 474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 420 IOLVVDVYFELTFLGVLGKFGITSAFSLVYFAELPTLVNRMAGVSTASRVSITIA 474
QY 475 APYFYLYGAVNMLPYIYMGSLVYIGITFLFPPEISGMLPETLEOMOKVKKFNSGK 534
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 480 APYFYLYGAVNMLPYIYMGSLVYIGITFLFPPEISGMLPETLEOMOKVKKFNSGK 534
QY 535 RDSMETEENPK 545
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 536 RDSLSKKGSPK 546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 7  
Q9VCA2 PRELIMINARY; PRT; 548 AA.

AC Q9VCA2; 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE ORC9 PROTEIN (GH21655P).  
GN ORC9 OR CG6331.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RA MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celliker S.E., Holt R.A., Evans C.A., Goeckyne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,





Db	157	KRPFTFASVIVLQITGCVLAANAPEYFSTTISRMTVGATTSCEVFLVAAYIALEMGSYSR- 215
Qy	229	IFSTLGVCTFFFAVGYMLPLFAFYFIRDMRMLLITVPGVLCVPMFIPESPMILSQR 288
Db	216	LFAGVAMQMFVSGVFNLTAGFAFYFIHDWRMQIAITPLPGLLFLCYMITIPESARMLMKG 275
Qy	289	RFRREDEDIQRKAANNNAVAVAFJDS-VEEELNPKQO-----KAFILDFPRTNIA 339
Db	276	RKDEPFLIEERKAKNKVEPNEIYEQLVDAVARKKQEDAMASOPATVFDLLRYPNLR 335
Qy	340	IMTIFSLMLMTSVGYALSLDAPNLHGDAVYLCNCFALSALIEPAYITAMLLIKTLTPRRY 399
Db	336	RKTLITFIDPMFVNSGVYYIGLSMNTNNGQNLQVPMNLISGPALIEIGYITLFFTLNRMGRS 395
Qy	400	IIAANLVFMGGVLLFTQLVPADYIFLISGLYMLCKRFGITSFAFSMLYFTELIVTLVRN 459
Db	396	ILCGTMVAAGISLTLATIVPBDMMVLVACMICKLMTISYGYITFYFSAEOPFTVVRN 455
Qy	460	AVGVSTSRSGSIAPFAFYVIGAVNRMLPIYWGSLTVLIGITFLEPPSIGMTLPETL 519
Db	456	GIGASSMARAROGILAPLKLKLGELWRPLPLITIGALSLTRGILSRLLPFTLNKMPETI 515
Qy	520	EQMKVQKWFRRSGKTRDSMETEE 542
Db	516	EDGENE-----GKKPAPQETAE 533
RESULT 9		
QYVCAS	9	PRELIMINARY; PRT; 567 AA.
AC	09VCAS3;	
DT	01-MAY-2000 (TREMBLREL, 13, Created)	
DT	01-MAY-2000 (TREMBLREL, 13, Last sequence update)	
DT	01-JUN-2001 (TREMBLREL, 17, Last annotation update)	
DE	CG13610 PROTEIN.	
CN	CG13610	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY;	
RC	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,	
RA	Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Ceahey S., Dahler C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodon K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Dunlun K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,	
RA	Jatall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Mekulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,	
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	

DE SD08136P.  
 GN CG13610.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN NCBI\_TaxID=7227;  
 RP SEQUENCE FROM N.A.  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuno J., Paclet J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
 RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY61615; AAL29163.1; .  
 SQ SEQUENCE 567 AA; 63234 MW; 3B10F3DBE19A0225 CRC64;

Query Match 28.6%; Score 815; DB 5; Length 567;

Best local Similarity 33.4%; Pred. No. 1.4e-51;  
 Matches 193; Conservative 112; Mismatches 192; Indels 80; Gaps 13;

QY 4 YDEVIATLGEWGPQRLIFLLSASIIIPNGFNGSVFLAGTPEHRCRVPDANLSSAMR 63  
 DB 3 YDEAIHLGDGRYQKITYFLICLTSPVAFHKLAVGLAKPDRFRCALP-----FE 54  
 QY 64 NNSVPLRLRDGREVP-----HSCSRYLATITANSALGLEGRVDL-----G 106  
 DB 55 NGSI-----YELPHTLMLNLSYPERNCRSY-----DVDTTEEXLNG 90  
 QY 107 QL-----EESCLDGEWSPQDYLTSTVTENNLVCEENMKVPLTSLFPGVLLGSFVSG 161  
 DB 91 SIPSSNETKTC--SSYIYDRSKYLSAVTEMLNLCGRDPMAATSDSLFMLGVLIGSTYFG 149  
 QY 162 QLSDRFGKRNLFATMAVOTGFSEFLQIFSISWEMFTVLF--VIVMGQISNYVAFLTGT 219  
 DB 150 QLSKRYGKRPILFASLVLOVLFVGLA--GVAPXEFTYTFALMVGATTSGLVLAAYVAM 207  
 QY 220 ELKSKVATIFSTLGCYCFEFAVGMLPLFAFYFTRDMRMILLATVGVLCVPLMWTPE 279  
 DB 208 EMVGSDRK-LYAGLFVMMFSGFMLTAVFAFYFVHDMRMQLIALTLGLLFMEFYWTLE 266  
 QY 280 SPRMLISORPREAEDITOKAKNNNTAVP-----AVTFDSVEELNPLKOOK----- 326  
 DB 267 SARMLLKGKRDCAIANNOAKARNKVEISDEALSELDEGENSEBAKAKLEQELDEG 326  
 QY 327 --AFLIDLFTRNIAIMTMSLLMLTSGYFALSIDAPNMGDAYLNCFLSALIEIPA 384  
 DB 327 PPPSVWDLFCYPNLRKRTLLIFLDMLVTSYGYGSLMWTSMLGCVLLNFIYSAGVEIPA 386  
 QY 385 YITAMLLLRTPRRITIAAVLFPGGCVLLFQIVPVYDYFLFSLGVMLGKRGITSAFMSML 444  
 DB 387 YIFLLTLNRMGRSILGCGLVMTGLSLATVILPQRMTVLIVACAMGLKATITASYGTV 446  
 QY 445 YVETAEIPTLVIRMAVGTSTASRVGSIIAPYFVYLGAVNRMPLPYIMGSLTVLIGIFT 504  
 DB 447 YIFSADQPTTYVRNVAAMASVARISSGMARPLNPLATITIKRPLPLIGSLTIVAGLLS 506  
 QY 505 LFEPSIGMTLPETLEOMOKVWFRSGKTRDSMTE 541  
 DB 507 LLLPETHNKPMLETIADGE-----RFGRKTKADVYLE 538

RESULT 11

Q96RU0 PRELIMINARY; PRT; 577 AA.

AC 096RU0;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)  
 DE ORGANIC CATION TRANSPORTER OKBI.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN NCBI\_TaxID=9606;  
 RP SEQUENCE FROM N.A.  
 RA Okabe M., Abe T.;  
 RT "Molecular identification of a novel organic cation transporter  
 OKBI.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF268892; AAK58593.1; .  
 SQ SEQUENCE 577 AA; 64584 MW; CIDDDF076C0488FD CRC64;

Query Match 27.9%; Score 794.5; DB 4; Length 577;

Best local Similarity 32.5%; Pred. No. 4.4e-50;  
 Matches 187; Conservative 117; Mismatches 222; Indels 49; Gaps 9;

QY 2 RDYEVIAFLGEWGPQRLIFLLSASIIIPNGFNGSVFLAGTPEHRCRVPDANLSSAMR 53  
 DB 4 RHEGTYDHYGHGRQRYQLYFICAFQNISCGIHLYASVFMGVTPHHCVRPGNVSQVVF 63  
 QY 54 -----DANLSSAMNNNSVPLRLRDGREVPHS--CSRYLATITANSALGLEGRVD 103  
 DB 64 HNHSNMSLEDTGALLSSGGKDYVTVOLONGEIMELSRCSRNR---ENTSLGYE----- 115  
 QY 104 DLGQLFOESCLDGEWSPQDYLTSTVTENNLVCEENMKVPLTSLFPGVLLGSFVSGQL 163  
 DB 116 YTSKKEPFQVQDYITDQMTKSTAYTQNNLYCDKRMKMLQPLFMEGVLLGSYTFGGF 175  
 QY 164 SDRFGKRNLFATMAVOTGFSEFLQIFSISWEMFTVLFVIVMGQISNYVAFLTGT 223  
 DB 176 SDRLGRRVYLMATSSSMFLGIAAFADVTFMARFPLAVASGLVVGCVYVMEFG 235  
 QY 224 KSVRIITFSLGCTEFPAVGMLPLFAFYFTRDM--RMILLATVGVLCVPLMWTPE 280  
 DB 236 MKSR-TWASVHHSFPAVGGLVALGLVLRFTWMLXOMLTSTVYFELIC--WVLPET 291  
 QY 281 PRMLISORPREAEDITOKAKNNNTA-----VPAVIDSVEELNPLKOKRAFLIDLER 334  
 DB 292 PFWLLSBGRKEQAKIVDMAKNNRASCKLSLSDLOGVSNSPTEVQKHNLSYLF 351  
 QY 335 TRNIAIMTMSLLMLTSGYFALSIDAPNMGDAYLNCFLSALIEIPATYITAMLLT 394  
 DB 352 NMSITRTLLVWMLMPTGSLFSGFSLNSLGNENYMLFLGLGVETIPAYFVFCIATDK 411  
 QY 395 LPRRTIIAALFPGGCVLLFQIVPVYDYFLFSLGVMLGKRGITSAFMSLTYETALPT 454  
 DB 412 VGRRTVLAAYSLFCSALACGVVAVIPQKHVILGVYTAAMVKFAIGAFGLIYLTAEIPT 471  
 QY 455 LVNMAVGTSTASRVGSIIAPYFVYLGAVNRMPLPYIMGSLTVLIGIFTLFPESLGMT 514  
 DB 472 IVSISLVAGSSMVCRLASIIAPFSVDLSIWIIFIPOLPFGTAMLLSGVLTIKLPELGR 531  
 QY 515 LPELTOMOKVWFRSGKTRDSMTEENPKVLIT 549  
 DB 532 LATTEBAAKL-----ESENESKSKILLT 556

RESULT 12

Q95TW6 PRELIMINARY; PRT; 563 AA.

AC Q95TW6;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)  
 DE GH27944P.  
 GN CG4630.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN NCBI\_TaxID=7227;  
 [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-V, CN BW SP;  
RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Lao G., Miranda A., Mungall C.J.,  
RA Munoz J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Gelinker S.,  
RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AY058475; AAL13704.1;  
SQ SEQUENCE 563 AA; 62138 MW; D3E2CB4096A195CB CRC64;

Query Match 27.5%; Score 781.5; DB 5; Length 563;  
Best Local Similarity 32.8%; Pred. No. 3.8e-49;  
Matches 190; Conservative 120; Mismatches 174; Indels 95; Gaps 17;

QY 3 DYDEVIAFLGEMGPQRLIFLLSASIIIPNGFNGSVVFLAGTPEHRCRPDANLSSA- 61  
DB 14 DFDLIVEIGERGRQRNRYLLICLPVLFPAANSLSYFTAGSPYRCYVECDKLVDAAE 73  
QY 62 ---WRNNSVP-LRLRDGEVPHSCSRYLATIANFSAL-GLBPGRDVDLGG----- 107  
DB 74 YGANWVSIAPGSMKRGHFTPTCERF---VANGDHLSSDPMSAMPIDQCAENFT 129  
QY 108 LEQESCLDGEFSQDYLS---TVVTENLVCEDN-WKVPITLSLFFYGVLLGSFVSQGL 163  
DB 130 TETERC-----NOFYVSSERTIVQMGLOCPENLWKLAEVGTIFHFGALVGTALSGYL 183  
QY 164 SDRFGKKNVLFATMAVQGFSLQIF-----SISWEMFVLF--VIYMGQISNVYV 213  
DB 184 ADRGGRKHI-----FLFCYFPMALITGVAQALSDIYSLFFALLNANVGTSGVYPL 233  
QY 214 AFLIGTEILGKSVRIITESTLGVCFPAVGYMLPLFAVFIIDWMLLATLTPGVLCPVL 273  
DB 234 AFLIGVEWVGPRKREM--SSIVLNFYAVGEALLGL-SVFLPDMQOLATLSVPLICVAY 291  
QY 274 WMFIPESPRMILISQRRFEADIIOKAKAMNTAVPAVIDSVBELMPLQOK----- 326  
DB 292 FVLVESYRMLLARNRBOAGVITIRRAKAVNRDI-----SVLWMSFKQOELDAETGQ 345  
QY 327 -----AFILDFRTNIAIMTISLMLTMSLVGVFALSILDPNLHGD 369  
DB 346 EDVEGGLHVKKDKIMLWAVEVARSHILMGRAVILLILNANVNIYYGSLNATISLGN 405  
QY 370 AYLWCFISALIEIPAYITAMLLTLPRLRIYAVALFWGGVLLFIQLVPVDYFISIG- 428  
DB 406 KYLNFAIVCLVEIPGYSIAMLWELRRFGRVALS-----GSLLLCSITCVASGEVTIGA 458  
QY 429 -----LVMLGKFGITSFMSLVYFAIEVPTLVNMAVGVTSASVSGSIIMPFYVYG 482  
DB 459 NMLVVTLEFLVKGKLGITTSFAVIYITFIDEMPTVIRSGGVMSIFARFGAMLAPFVPLA 518  
QY 483 AYNRMLEPIYVGSLLVIGITLFFPESLGTLPETLEQ 521  
DB 519 SYDPLPLLEGLTSLVAGLLSLLPETFNKRLPDYEE 557

RESULT 13  
Q9V6L6 PRELIMINARY; PRT; 561 AA.  
AC Q9V6L6;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE CG4630 PROTEIN.  
GN CG4630.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Gelinker S.E., Holt R.A., Evans C.A., Goeyne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zheng Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Bruttier P.,  
RA Burtils K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Moadary C., Morris J., McShrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler E., Shen H.,  
RA Shue B.C., Siden-Klamos I., Stapleton M., Skupski M.P., Smith T.,  
RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
DR EMBL: AEO03819; AAF56407.1; --  
DR Flybase: FBgn0033809; CG4630.  
DR InterPro: IPR003662; sub. transporter.  
DR Pfam: PF00083; sugar\_tr; 1.  
KW Transmembrane.  
SQ SEQUENCE 561 AA; 62041 MW; 573643481595788F CRC64;

Query Match 27.2%; Score 773.5; DB 5; Length 561;  
Best Local Similarity 32.8%; Pred. No. 1.5e-48;  
Matches 189; Conservative 118; Mismatches 174; Indels 95; Gaps 17;

QY 3 DYDEVIAFLGEMGPQRLIFLLSASIIIPNGFNGSVVFLAGTPEHRCRPDANLSSA- 61  
DB 14 DFDLIVEIGERGRQRNRYLLICLPVLFPAANSLSYFTAGSPYRCYVECDKLVDAAE 73  
QY 62 ---WRNNSVP-LRLRDGEVPHSCSRYLATIANFSAL-GLBPGRDVDLGG----- 107  
DB 74 YGANWVSIAPGSMKRGHFTPTCERF---VANGDHLSSDPMSAMPIDQCAENFT 129  
QY 108 LEQESCLDGEFSQDYLS---TVVTENLVCEDN-WKVPITLSLFFYGVLLGSFVSQGL 163  
DB 130 TETERC-----NOFYVSSERTIVQMGLOCPENLWKLAEVGTIFHFGALVGTALSGYL 183  
QY 164 SDRFGKKNVLFATMAVQGFSLQIF-----SISWEMFVLF--VIYMGQISNVYV 213  
DB 184 ADRGGRKHI-----FLFCYFPMALITGVAQALSDIYSLFFALLNANVGTSGVYPL 233  
QY 214 AFLIGTEILGKSVRIITESTLGVCFPAVGYMLPLFAVFIIDWMLLATLTPGVLCPVL 273  
DB 234 AFLIGVEWVGPRKREM--SSIVLNFYAVGEALLGL-SVFLPDMQOLATLSVPLICVAY 291

```

QY 274 MWIPESPRMLISORPREAEDIIOKAKMNTNAPVAVIFDSVEELNPLKOOK----- 326
Db 292 FWLPESVRMLLARRRBOGVIIIRAKAVNRDI-----SVLMSEFOQOELDAETQ 345
QY 327 -----AFILDFETRNIAIMTISLLMLTSSVGFALSALDAPNLHGD 369
Db 346 EDDVEGGLHWKKDKIMLAVKEVARSHILMGRAVILLILMAVNAIVYGLSLNATSLSGN 405
QY 370 AYLNCFLSALIEIPAYITAMLLTLPRRYIAAIVFWGCVLLFIOLVVDYIFLSIG- 428
Db 406 KYLNFALVCAIEIPGYSIAMFLFRFRGRVALS-----GSLILCSITCVAISGFVTLGA 458
QY 429 -----LVMLGKFGTTSFASMLVYETAELVPTLRNMAVGTSTASRVGSIIAFYFYL 482
Db 459 NMLVVTFLVGLKLTSSFAVITYFTLEKMTVIRSGCVGMSIFARFGAMLAFVFLA 518
QY 483 AYNRLPYVMSGLTVLIGITLFFPESIGMTLPET 518
Db 519 SYDPLPLFLFTLSLVAGLLSLLPETFNKLPDT 554

```

```

RESULT 14
QY0539 PRELIMINARY; PRT; 568 AA.
AC Q9U539;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ORGANIC CATION TRANSPORTER 1 (F52F12.1B PROTEIN).
GN OCT-1 OR F52F12.1B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99227113; PubMed=10209228;
RA Wu X., Fei Y.J., Huang W., Chaney C., Leibach F.H., Ganapathy V.;
RT "Identity of the F52F12.1 gene product in Caenorhabditis elegans as an
RT organic cation transporter.";
RL Blochim. Biophys. Acta 1418:239-244(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; AF110415; AAF21932.1; -.
DR EMBL; Z83228; CAC70093.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 568 AA; 63505 MW; C0A3E73851F44056 CRC64;

```

```

Query Match 26.5%; Score 754.5; DB 5; Length 568;
Best Local Similarity 30.2%; Pred. No. 3.6e-47;
Matches 169; Conservative 119; Mismatches 236; Indels 35; Gaps 9;
QY 3 DYDEVIAFLGEMGPQRLIFFLS-ASIIPNGFGMSVFLAGTPRRCRVDAANLSSA 61
Db 8 DDFVLEQVGNVGTQIVFFELIPTSLPSAFAFNIPFVGNPNPHCHIEGKEYLRP 67
QY 62 WRNNSVPLRLRQGRVPHSSKRLATINFSALGIEPGRVDVLQDEESCLDGEFSQ 121

```

```

Db 68 LFNDOIL-----SKOYNETOINVFRATISAP-VDTYSDRISLVPQNGMDYDN 116
QY 122 DYSLTVVEMNLVCEDNKVPPLTSLFVGLVSFGVSGOLSDPRGRKNVFAFMAYOT 181
Db 117 STYLSLVTEFNLVCDQAMIEISTTSFYVGSFIGNCLGYADFGRRRSFVLLYLI 176
QY 182 GFSFLIOFISISWEMFTVLEIVMGQISNVVAIFLIGTEILKSVRIIFSTLGVCTPFAV 241
Db 177 VCGTASSPAKDIESFILLRPFGLAPALFQIPFTICMEFMNSGR-IFSGIMTSLFPGA 235
QY 242 GYVLLPLFAFVIFDMKMLLALTVPGVLCVPLMWFIPESPRMLISORPREAEDIIOKAA 301
Db 236 AMALLGVVAMFIRKROLFFPCNAPPAFYIIYFFLPESPRMSVSYGAADKKOIKIA 295
QY 302 KMN---NTAVPAVIFDSVEELNPLKOOKAF-----ILDFETRNIAIMTISLLMLTSS 353
Db 296 KMGKGSNVDDVDELV-DSMKNHQAAMEKETKSHNTDLPKPNLRKLLITYIIVMNA 354
QY 354 VGYFALSALDAPNLHGDAYLNCFLSALIEIPAYITAMLLTLPRRYIAAIVFWGCVLL 413
Db 355 IIVNGTLVNSNLPVDYMSFIINGAVELPGYFVVPPLLCAGRRWTLAATMIVCGIGCV 414
QY 414 FIOLVVDYIFLSIGVLMGKFGITSAFMSLYETAELVPTLRNMAVGTSTASRVGSI 473
Db 415 SAMFMPDGYPMVAASAFIGKFGVSGFAVITYFAGELPVPYRAIGKMSVWAGSGLL 474
QY 474 IAPFVYLGAYNMPLPYVMSGLTVLIGITLFFPESIGMTLPETLEQOKVWFRSGRK 533
Db 475 LAHVIVLGVIVITLPLLMGLMALNSGILTFPLPETLGAPLPMTIEDANF-----GKK 529
QY 534 -----TRDSMETEENP 544
Db 530 PEPDSGMFTQAKKRRESQ 548

```

```

RESULT 15
QY02270 PRELIMINARY; PRT; 576 AA.
AC Q02270;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE F52F12.1A PROTEIN.
GN F52F12.1A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; Z83228; CAB05732.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 576 AA; 64493 MW; F494E94A7EBC0B1 CRC64;

```

```

Query Match 26.5%; Score 754.5; DB 5; Length 576;
Best Local Similarity 30.2%; Pred. No. 3.7e-47;
Matches 169; Conservative 119; Mismatches 236; Indels 35; Gaps 9;
QY 3 DYDEVIAFLGEMGPQRLIFFLS-ASIIPNGFGMSVFLAGTPRRCRVDAANLSSA 61

```

```
Db      16 DEDFVLEQVNGTGYQIVFEFFIICLPTSLPSAFSAFNIPIFVGNPNPHTCHIPEGEKEYLRP 75
QY      62 WNNNSVPLRLRDGREVPSCSRYLATIANPSALCLEPGRDVIDGLQLEQESCLDGEFSQ 121
Db      76 LINDTOIL-----SKOYNETOQINVPFAFTSAP-VDITYSDRISLVPCQNGWDYDN 124
QY      122 DYLSTVVTENMLVCEDMWKVPLTSLFPEVGLLGSFVSGQLSDPRGRKNLFTMAVQT 181
Db      125 STYLDLSTVEFNLVCDQAMIEISTTSFYVGSFIGNCLFGYVADKFGRRSFVILYLI 184
QY      182 GSFLOIFSISEMFTVFIYVGMQISNYVAFTLGTEILGKSVRIESTLGVCTFEFV 241
Db      185 VCGTASSFAKIDIESFIILRFTGLAFPALFOIPFIICMEFMGNSGR-IFSGIMTSLFEFGA 243
QY      242 GYMLPLFAFYIRDRMULLLALTVPGVLCVPLMWFIPESPRMLISQRRFREADILQKA 301
Db      244 AMALLGVVAMFIRRKQTLFEFCNAPFAFYIITYFFLPESPRWSVSVGKWADAKQLKRIA 303
QY      302 KMN---NTAVPAVIFDSVYEELNPLKQKAF-----IDLFRTNIAIMTISLMLMLTS 353
Db      304 KMGKSNVDVDELV-DSMKNHQNAAEEKETKRSNVTDLFEKTPMLRRKTLIVYIWMNA 362
QY      354 VGYFRLSDAENLHGDALNCLFSLALIEIPAYITAMLLRLTPRRYIIAAVLEWGGVLL 413
Db      363 IYNGILTVSNLPPVDDWISFLINGAVELPGYFVWPLLOACAGRRMTLAATMIVCGIGCV 422
QY      414 FIOLPVDVYFSLIGLWMLGKFGITSFMSLYFTAELIPTLVNMAVGVSTASRVGSI 473
Db      423 SAMFMPOGYPMLVASASAFIGKFGVSGFAVYIIFAGELIYTVARAIGMSSMVAAGSGL 482
QY      474 IAPYFVYIGAYNNRMPLIYVWGLTVLIGITLFFPESLGMTLPETLEOMOKVKWFRSGK 533
Db      483 LAPHIVNLGKIVKILPILMIGIMALSAGILTFELPETLGAPLPMTIEDAENE-----GKK 537
QY      534 -----TRDSMETENP 544
Db      538 PEPDSGMFTQAKAKKRESOP 556
```

Search completed: July 17, 2002, 02:14:09  
Job time: 655 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2002, 19:36:09 ; Search time 7127.59 Seconds  
(without alignments)  
6268.340 Million cell updates/sec

Title: us-09-521-195b-2

Perfect score: 2135  
Sequence: 1 ccccgctcgcgcccaat.....aatactatccaataaataat 2135

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb.ba:\*  
2: gb.hg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*  
29: em.vi:\*  
30: em.htg\_hum:\*  
31: em.htg\_inv:\*  
32: em.htg\_other:\*  
33: em.htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

Result No.	Score	Query Match	Length	ID	Description
1	2135	100.0	2135	9	AB007448 Homo sapi
2	2130.2	99.8	2135	9	AB007448 HSPTRMP
3	1310.6	61.4	2083	10	AB016257 Mus muscu
4	1310.6	61.4	2277	10	BC010590 Mus muscu
5	1293	60.6	2258	10	AF169831 Rattus no
6	1066.4	49.9	1831	9	AB015050 Homo sapi
7	1065.8	49.9	3252	9	AF057164 Homo sapi
8	1062.6	49.8	2811	9	BC012325 Homo sapi
9	994	46.6	1881	10	AF111425 Mus muscu
10	994	46.6	1888	10	AB015800 Mus muscu
11	992.4	46.5	2621	10	AF110417 Mus muscu
12	983.6	46.1	3007	10	RNAJ1933 Rattus no
13	982.8	46.0	3037	10	AF110416 Rattus no
14	978.8	45.8	3038	10	AB017260 Rattus no
15	905.6	42.4	2297	10	AB018436 Mus muscu
16	539	25.2	3081	9	HS181760 Homo sapien
17	539	25.2	89174	9	AC008599 Homo sapien
18	539	25.2	146797	9	AC034220 Homo sapi
19	539	25.2	191397	2	AC023861 Homo sapi
20	505	23.7	700	6	AX183016 Sequence
21	410	19.2	122557	2	AC011429 Homo sapi
22	410	19.2	191397	2	AC023861 Homo sapi
23	404	18.9	3848	9	HSAC000954 Homo sapien
24	404	18.9	3848	9	HUM17DC32 Homo sapien
25	363.8	17.0	700	6	AX182500 Sequence
26	363.8	17.0	25871	9	AB016625 Homo sapi
27	328.4	15.4	191732	2	AL596182 Mus muscu
28	321.4	15.1	93409	2	AC073294 Mus muscu
29	321.4	15.1	154256	2	AC073295 Mus muscu
30	321.4	15.1	171274	2	AL596444 Mus muscu
31	315.6	14.8	107850	2	AC035148 Homo sapi
32	314	14.7	700	6	AX182552 Sequence
33	297	13.9	700	6	AX182553 Sequence
34	234	11.0	700	6	AX183015 Sequence
35	217.6	10.2	700	6	AX182936 Sequence
36	197	9.2	2106	10	RNOCT2R Rattus norv
37	195.6	9.2	2152	10	D83044 Rattus norv
38	195.2	9.1	1896	6	A48159 Sequence 6
39	193.6	9.1	2257	9	X98333 H.sapiens m
40	193	9.0	2078	10	BC015250 Mus muscu
41	193	9.0	2164	10	AJ006036 Mus muscu
42	190.8	8.9	2132	10	X98334 R.norvegicu
43	189.6	8.9	2064	3	DMORCT2 Drosophila
44	189.6	8.9	2564	3	AY058437 Drosophila
45	189.6	8.9	156905	2	AC020006 Drosophila

## ALIGNMENTS

Result No.	Score	Query Match	Length	ID	Description
1	2135	100.0	2135	9	AB007448 Homo sapi
2	2130.2	99.8	2135	9	AB007448 HSPTRMP
3	1310.6	61.4	2083	10	AB016257 Mus muscu
4	1310.6	61.4	2277	10	BC010590 Mus muscu
5	1293	60.6	2258	10	AF169831 Rattus no
6	1066.4	49.9	1831	9	AB015050 Homo sapi
7	1065.8	49.9	3252	9	AF057164 Homo sapi
8	1062.6	49.8	2811	9	BC012325 Homo sapi
9	994	46.6	1881	10	AF111425 Mus muscu
10	994	46.6	1888	10	AB015800 Mus muscu
11	992.4	46.5	2621	10	AF110417 Mus muscu
12	983.6	46.1	3007	10	RNAJ1933 Rattus no
13	982.8	46.0	3037	10	AF110416 Rattus no
14	978.8	45.8	3038	10	AB017260 Rattus no
15	905.6	42.4	2297	10	AB018436 Mus muscu
16	539	25.2	3081	9	HS181760 Homo sapien
17	539	25.2	89174	9	AC008599 Homo sapien
18	539	25.2	146797	9	AC034220 Homo sapi
19	539	25.2	191397	2	AC023861 Homo sapi
20	505	23.7	700	6	AX183016 Sequence
21	410	19.2	122557	2	AC011429 Homo sapi
22	410	19.2	191397	2	AC023861 Homo sapi
23	404	18.9	3848	9	HSAC000954 Homo sapien
24	404	18.9	3848	9	HUM17DC32 Homo sapien
25	363.8	17.0	700	6	AX182500 Sequence
26	363.8	17.0	25871	9	AB016625 Homo sapi
27	328.4	15.4	191732	2	AL596182 Mus muscu
28	321.4	15.1	93409	2	AC073294 Mus muscu
29	321.4	15.1	154256	2	AC073295 Mus muscu
30	321.4	15.1	171274	2	AL596444 Mus muscu
31	315.6	14.8	107850	2	AC035148 Homo sapi
32	314	14.7	700	6	AX182552 Sequence
33	297	13.9	700	6	AX182553 Sequence
34	234	11.0	700	6	AX183015 Sequence
35	217.6	10.2	700	6	AX182936 Sequence
36	197	9.2	2106	10	RNOCT2R Rattus norv
37	195.6	9.2	2152	10	D83044 Rattus norv
38	195.2	9.1	1896	6	A48159 Sequence 6
39	193.6	9.1	2257	9	X98333 H.sapiens m
40	193	9.0	2078	10	BC015250 Mus muscu
41	193	9.0	2164	10	AJ006036 Mus muscu
42	190.8	8.9	2132	10	X98334 R.norvegicu
43	189.6	8.9	2064	3	DMORCT2 Drosophila
44	189.6	8.9	2564	3	AY058437 Drosophila
45	189.6	8.9	156905	2	AC020006 Drosophila

JOURNAL FEBS Lett. 419 (1), 107-111 (1997)  
 MEDLINE 98086199  
 FEATURES  
 source Location/Qualifiers  
 1. 2135  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /tissue\_type="liver"  
 /dev\_stage="fetal"  
 /note="inbetween D5S642 and D5S643"  
 147. .1802  
 /gene="OCTN1"  
 147. .1802  
 /name="OCTN1"  
 /note="Organic cation transporter OCTN1 (fls631)"  
 /codon\_start=1  
 /product="OCTN1"  
 /protein\_id="AAA23356.1"  
 /db\_xref="GI:2605501"  
 /translation="MRDDEVIAFLGMEPFQRLIFLLSNLIPNGMNVFLAG  
 VPEHRCRVDPDANLSAMRNNSVPLRLDGRVPHSCSRKLATIANFSLGLEPGRD  
 VDLQLEQESCLDMEFSQDYVLTSTVYEMNLVCEDMNKVPLTSLFFVGLGSPFS  
 GQSLDRFGRKVLFPATMAVOTGFSLQIFSLSEMFTVLVIVCMGQISNVYAFILG  
 TEIGKSVRIEISLGYCTFPVAVGYMLPLPARIIRMRMLLALIVPVLCYPLMFE  
 IPESPRLISORRREADIIQAKANNRAVPVITDSVELNPLKQNAFTIDLER  
 TRNAYIMTISLMLMLSVGFALSLDAPLHGDAYLNCFLSLLEIPAYITAMLLR  
 RTPLRYITIAVLVWGGVLEFLQVLPVDYVFLSIGLVMKRGITSAFSLVYTFAL  
 LYPPLVRMAVGVSTASRVGSIIPAPYFVVLGAVNRMLPYIVMSLTVLIGITLFFP  
 LSISMLTPELLEOMOKVWFERSKTRDSMETENPKVLITAF"

gene  
 CDS  
 147. .1802  
 /gene="OCTN1"  
 147. .1802  
 /name="OCTN1"  
 /note="Organic cation transporter OCTN1 (fls631)"  
 /codon\_start=1  
 /product="OCTN1"  
 /protein\_id="AAA23356.1"  
 /db\_xref="GI:2605501"  
 /translation="MRDDEVIAFLGMEPFQRLIFLLSNLIPNGMNVFLAG  
 VPEHRCRVDPDANLSAMRNNSVPLRLDGRVPHSCSRKLATIANFSLGLEPGRD  
 VDLQLEQESCLDMEFSQDYVLTSTVYEMNLVCEDMNKVPLTSLFFVGLGSPFS  
 GQSLDRFGRKVLFPATMAVOTGFSLQIFSLSEMFTVLVIVCMGQISNVYAFILG  
 TEIGKSVRIEISLGYCTFPVAVGYMLPLPARIIRMRMLLALIVPVLCYPLMFE  
 IPESPRLISORRREADIIQAKANNRAVPVITDSVELNPLKQNAFTIDLER  
 TRNAYIMTISLMLMLSVGFALSLDAPLHGDAYLNCFLSLLEIPAYITAMLLR  
 RTPLRYITIAVLVWGGVLEFLQVLPVDYVFLSIGLVMKRGITSAFSLVYTFAL  
 LYPPLVRMAVGVSTASRVGSIIPAPYFVVLGAVNRMLPYIVMSLTVLIGITLFFP  
 LSISMLTPELLEOMOKVWFERSKTRDSMETENPKVLITAF"

BASE COUNT 499 a 547 c 530 g 559 t  
 ORIGIN

Query Match 100.0%; Score 2135; DB 9; Length 2135;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cccggcctcgcccccatttctcaagcgtcgtcccccgggaagcttcaatc 60  
 Db 1 cccggcctcgcccccatttctcaagcgtcgtcccccgggaagcttcaatc 60  
 Qy 61 cttgggagcgcccaagctacaagcaactgtccctgagaagcgtctcaacgctagt 120  
 Db 61 cttgggagcgcccaagctacaagcaactgtccctgagaagcgtctcaacgctagt 120  
 Qy 121 caagtttcgagcgagctgggaagcatgcggaactacgagcaagctgacgcttcctg 180  
 Db 121 caagtttcgagcgagctgggaagcatgcggaactacgagcaagctgacgcttcctg 180  
 Qy 181 gcgagtggggccctccagcagcctcatctctcctgctcgaagcagatatacccca 240  
 Db 181 gcgagtggggccctccagcagcctcatctctcctgctcgaagcagatatacccca 240  
 Qy 241 atggctcaatgtatgtcagctgtctcctgcgaggagaccccggaacagcgtctcga 300  
 Db 241 atggctcaatgtatgtcagctgtctcctgcgaggagaccccggaacagcgtctcga 300  
 Qy 301 tgcggagcgcccggaacactgagcagcgctcgcgcaacaacagtgtccgctcgagctgc 360  
 Db 301 tgcggagcgcccggaacactgagcagcgctcgcgcaacaacagtgtccgctcgagctgc 360  
 Qy 361 gggagcgcccggaagctgcccacaagctgcaagcgctacgagctcgccacatcgccaact 420  
 Db 361 gggagcgcccggaagctgcccacaagctgcaagcgctacgagctcgccacatcgccaact 420  
 Qy 421 tctcgagcctcgagctggaagccggggcgagcgtggaacctgaggcaagctggaagcagaga 480  
 Db 421 tctcgagcctcgagctggaagccggggcgagcgtggaacctgaggcaagctggaagcagaga 480  
 Qy 481 gctgcctgagctgagtgagtgatcaagccagaagcttcaacctgtccacgctctggaacgagt 540  
 Db 481 gctgcctgagctgagtgagtgatcaagccagaagcttcaacctgtctccacgctctggaacgagt 540

Qy 541 ggaatctgtgtgtgagagacaactcgtgaaggtgcccctcaaccctccctgttctgtag 600  
 Db 541 ggaatctgtgtgtgagagacaactcgtgaaggtgcccctcaaccctccctgttctgtag 600  
 Qy 601 gctgtcctcctgctcctctgtccgggacgctgtcagaagcttggaggagaagaag 660  
 Db 601 gctgtcctcctgctcctctgtccgggacgctgtcagaagcttggaggagaagaag 660  
 Qy 661 tctcttcgacaacatgctgtgtacagacgtgctcaagcttcctcgtgagaatttccatca 720  
 Db 661 tctcttcgacaacatgctgtgtacagacgtgctcaagcttcctcgtgagaatttccatca 720  
 Qy 721 gctggagagatgtaacgtgtattgttcgtgtgagcagtgaggccagatctccaactatg 780  
 Db 721 gctggagagatgtaacgtgtattgttcgtgtgagcagtgaggccagatctccaactatg 780  
 Qy 781 tggtagccttcaactagagacaagaatctcttgcaagtagtctgattatattctctca 840  
 Db 781 tggtagccttcaactagagacaagaatctcttgcaagtagtctgattatattctctca 840  
 Qy 841 catlaagagctgagcagatgtctgtcgctggcgctgagcagtgagtgctgtgtgcttact 900  
 Db 841 catlaagagctgagcagatgtctgtcgctggcgctgagcagtgagtgctgtgtgcttact 900  
 Qy 901 tcaatcagaagctgagcagatgtctgtcgctggcgctgagcagtgagtgctgtgtgcttact 960  
 Db 901 tcaatcagaagctgagcagatgtctgtcgctggcgctgagcagtgagtgctgtgtgcttact 960  
 Qy 961 cgcctgtgtgttcaatctcctgaatctcccgagctgagctgatatcccaagaagattagag 1020  
 Db 961 cgcctgtgtgttcaatctcctgaatctcccgagctgagctgatatcccaagaagattagag 1020  
 Qy 1021 aggcctgaagatcatcatccaaaagctgcgcaaaaatggaacaacagctgtcaacagcagtga 1080  
 Db 1021 aggcctgaagatcatcatccaaaagctgcgcaaaaatggaacaacagctgtcaacagcagtga 1080  
 Qy 1081 tattgattctgtgagagagctaaacccttgagcagcagaagaagcttcatctctgagac 1140  
 Db 1081 tattgattctgtgagagagctaaacccttgagcagcagaagaagcttcatctctgagac 1140  
 Qy 1141 tgttcaggactcgggaatatgtccataaagcaattatgtcttgcgtgctgagatgtctga 1200  
 Db 1141 tgttcaggactcgggaatatgtccataaagcaattatgtcttgcgtgctgagatgtctga 1200  
 Qy 1201 cctcagtggttactctgtcctcgtctcgtgatgtccttaattaaatgaaatgtgcttacc 1260  
 Db 1201 cctcagtggttactctgtcctcgtctcgtgatgtccttaattaaatgaaatgtgcttacc 1260  
 Qy 1261 tgaactgttctcctcgtccttgatgaaatcccaagcttacaatcaagcgtgtgctgtat 1320  
 Db 1261 tgaactgttctcctcgtccttgatgaaatcccaagcttacaatcaagcgtgtgctgtat 1320  
 Qy 1321 tgggaacgctgcccaggggttatcatatgctgcagtaetgtctcggggaagagtgatgc 1380  
 Db 1321 tgggaacgctgcccaggggttatcatatgctgcagtaetgtctcggggaagagtgatgc 1380  
 Qy 1381 tctcttcaatcaactgtgacctgtggaattatcaatcttcaatcttcaatcttcaatcttca 1440  
 Db 1381 tctcttcaatcaactgtgacctgtggaattatcaatcttcaatcttcaatcttcaatcttca 1440  
 Qy 1441 tgggaaatattgggaatcaactctgcttctccatgacgtgtatgtctcctgtagacttc 1500  
 Db 1441 tgggaaatattgggaatcaactctgcttctccatgacgtgtatgtctcctgtagacttc 1500  
 Qy 1501 acccaaccctgttcaaggaaatcgtgcggtggggtacactcaacggcctcccaagagtggca 1560  
 Db 1501 acccaaccctgttcaaggaaatcgtgcggtggggtacactcaacggcctcccaagagtggca 1560  
 Qy 1561 gcatcattgccccctacttgttlaactcgtgtcgttcaacaagaatgtgtgcctcaatcag 1620  
 Db 1561 gcatcattgccccctacttgttlaactcgtgtcgttcaacaagaatgtgtgcctcaatcag 1620  
 Qy 1621 tcatgtgtatgtcagctgtcctgattggaatccttcaacccttcttccctgaaagttag 1680



|||||  
Db 1621 TCATGGGAGTCTACTGTCTCTGATTTGGAATCTTCACCCCTTTTTCCTCCGGAAGTTGG 1680  
Qy 1681 gaatgactctccagaaaccttagagacagatgcagaaagtgaatggcttagatcagcggga 1740  
Db 1681 GAATGACTCTCCAGAAACCTTAGAGCAGATGCAGAAAGTGAATGGTTCAGATCTGGGA 1740  
Qy 1741 aaaaaaacagagactcaatgagacagaaagaaatcccaaggttctatactgcatctc 1800  
Db 1741 AAAAAACAAGAGACTCAATGGAGACAGAAATCCCAAGGTTCTAATAACTGCATTCCT 1800  
Qy 1801 gaaaaaatctaccoccatcttgatgagtgaaagacagaaataaagaccctgtagagaa 1860  
Db 1801 GAAAAAATATCTACCCCATTTGGTGAAGTGAAGAAACGAAAAAATAACACCCCTGGAGAA 1860  
Qy 1861 attcggtgtccacactgaatgagactgactgtaacgaattgacaccaaalgaaacctgtct 1920  
Db 1861 ATTGGTGTCTCCACTGGAATGAGACTGCTGATGCAATTCGACCAAAATGAACTTCGCT 1920  
Qy 1921 atcaagaatgctcgtcatagacagtaaacctcgtgagatcttcacagataatgctcgtct 1980  
Db 1921 ATCAAGAATGCTCGTATACAGTAAGTAACCTGATGATCTTCACAGATTAATGCTTCGCT 1980  
Qy 1981 ttacaacaacacatcttcataagagctcctactcaatlaatcaatgaatgagatgct 2040  
Db 1981 TTACAACAACACCATTTCTTAGAGAGTCTCCTACTCAATTAATCAATGAATGATTCGCT 2040  
Qy 2041 aagatgtctgaacaacatgcttagtaagcagactggttaaaatcataataagatacactc 2100  
Db 2041 AAGATGTCTTGAAACATGTAGTCAAGAGACTGTAATAATCATATAAGATTAACACTC 2100  
Qy 2101 atttccaatcatcaaatatactatccaataaanaat 2135  
Db 2101 ATTTCCAATCATACAAATATGATTCAAATTAATAAAT 2135

RESULT 2  
HSPIMP 2166 bp mRNA linear PRI 06-JAN-2001  
LOCUS HSPIMP  
DEFINITION H. sapiens mRNA for putative integral membrane transport protein.  
ACCESSION Y09881.1 GI:12053560  
VERSION Y09881.1  
KEYWORDS integral membrane protein; transport protein; UTR2h gene.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 2166)  
AUTHORS Spritzemberger, F., Gruendemann, D. and Schoemig, E.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2166)  
AUTHORS Gruendemann, D.  
TITLE Direct Submission  
JOURNAL Submitted (05-DEC-1996) D. Gruendemann, University Of Heidelberg,  
Department Of Pharmacology, Im Neuenheimer Feld 366, 69120  
Heidelberg, FRG

FEATURES  
source  
1. Location/Qualifiers  
1. 2166  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Caki-1"  
159. 1814  
/gene="UT2h"  
159. 1814  
/note="putative"  
/codon\_start=1  
/product="integral membrane transport protein"  
/protein\_id="CAA71007.1"  
/db\_xref="gi:12053561"  
/translation="MR0YDEVIAPIGEMSPRORLIFPLISIIIPNGNGSVYFIAG  
TPEHRCVPDANLSSAARNNSVPLRLDGEKREPHSCRSKYLATLATSALGLEPGRD  
VDLGQLEQESCIDGWEFSODVYLSTVTEWNLVGEDNWKVPLTTSLEFVGLISFVS

GOLDFEGRKNVLEFATMAVOTFSELOJFISISMETVLEFVIVMGQISNVVAFILG  
TEILGKSVILFESTIGCTFPVAGGMLPLFAVYIRRMMLLALTPVGLCVLMMF  
IPEPRMLISORPREAEDIIOKAKAMNNINAVPVIDSVSEINPLKQOAFILDLR  
TRNIAITMISLLIMLMSVSPALSIDAPMLHDHAIKNCPLAIEIPVITAMILL  
RLPFRYITIAVLEWGGVLLFQLVLPVDYFLSIGLMLCKRSTTSAFSLYFTAE  
LYPLVIRMMAGVSTASRVSIIAPVFIYGAANNRMIPYVMKSLTVLIGILFLFRP  
ESIGMTLEPETEOMOKVMFRSGKTRDRSMETENPKVILITAF"  
BASE COUNT 509 a 556 c 535 g 566 t  
ORIGIN

Query Match 99.8%; Score 2130.2; DB 9; Length 2166;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2132; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ccccggtctcgccgcccaattctcaacagcctgctgtcccccgggaagcttcaatc 60  
Db 13 CCCCGGCTTCGCGCCCAATTTCTTAACAGCCCTGCTCCCGGGAACGTTCTAACATC 72  
Qy 61 ctggggagagcccgactacaagacactgctcttgagaaagctgtcatcacccgtagt 120  
Db 73 CTGGGGAGCGCCCGAGCTACAAAGACACTGCTTGGAACGCTGTATACCCGTAGTTG 132  
Qy 121 caagtttcggagcgagtgagtgagacatgcygagactacgacgagtgatcgctctc 180  
Db 133 CAAGTTTCGAGCGGCGAGTGGGAACATGCGGGAGTACGACGAGTGTGCTTCCTGG 192  
Qy 181 gcaagtgaggggccccctcccgagcgcctatctctctctgtctcaagcgcaatcatccca 240  
Db 193 GCAAGTGGGGGCCCTTCACAGCGCTCATCTCTCTGTCTCAGCGCACACATCATCCCA 252  
Qy 241 atggcttcaatgtgtatgtagtgcgtgtctctgcyggggagcccccgagcagcgtgtc 300  
Db 253 ATGGCTTCAATGGTATGTCTAGTGTGTCTGTGGCGGGGAGACCCGGAGCAACGGTGTGAG 312  
Qy 301 tgcggagcgcgcggaactgtagcagcgctgycgcaacaacagtgctccgctgcygctgc 360  
Db 313 TGCGGAGCGCGCGAACCCTGAGCAGCGCCCTGGCGCAACAACATGTGCCGCTGCGCTCG 372  
Qy 361 ggaagcgccgagcgagtgcccccaacagctgagcgcgtacagcgctcgccacacatcgcaact 420  
Db 373 GGAAGCGCCCGAGGTGCCCCACAGCTGCAAGCGGCTACCGGCTGCCACCATGTGCCAACT 432  
Qy 421 tctcgcgctcgagcgtbtagacgcgggycgagcttgagaccttgaggcagcttgaggacagaga 480  
Db 433 TCTCGGCGCTCGGGCTGTGAGACCGGGGCGGAGCACTGGGGCAGCTGTGAGACAGAGA 492  
Qy 481 gctgctgtagtgcgtgagtgtaagccaagacgtctactcgttccacocgtctgacccgagt 540  
Db 493 GCTGCTGATGGCTGGAGGATTCAGCCAGGACGCTCACTGTCCACCGCTGTGACCGAGAT 552  
Qy 541 ggaatctggtgtgtgagagacaactggagagtgccccccacacactccctgttctcgtag 600  
Db 553 GGAATCTGGTGTGTGAGAGACAACGTGAAGGTGCCCTTACACACTCCCTCTTCTTGTAG 612  
Qy 601 gcgtgctcctcgctcctctgctgtccggcgagctgtctagacagagtttgagcaagaagaag 660  
Db 613 GCGTGTCTCTGCTCTCTGTGTGTCGGGCGAGCTGTGACAGAGTTGGCAGGAAGAAGC 672  
Qy 661 ttctcttcgcaacatgctgctgtacagactggtctcagcttccctcgcaagatlttctcatca 720  
Db 673 TTCTCTTCGCAACCATGGCTGTACAGACTGGCTTCACTTCGCAATTTTCTTCATCA 732  
Qy 721 gctggagagtgtaactgtgtatctgtcatcgtggcgatgggacagatcccaactatg 780  
Db 733 GCTGGAGATGTTCACTGTGTATTGTTCATCGTGGGCAATGGGCAATGCCAATCAGTAG 792  
Qy 781 tgttagccttcaactaggaacagaatctcttgcaagtcagtgatcgtatataatctctca 840  
Db 793 TGTAGCCTTATCACTAGGAACAGAAATTCCTGGCAAGTCAAGTTCGATATATATTCCTTA 852  
Qy 841 cattagaggtgtgacatlttttgcagttgctatatgcgcgtgcgcactgttggcttact 900

Db 853 CATTAGAGCTGTGCACATTTTTCGAGTTTGCAATATATGCTGCTGCCACTGTTTCTACT 912

Qy 901 tcatcagagactgaggagatgctgctgctgaggctgaagtgctgagatgctgctgctgctc 960

Db 913 TCATCAGAGACTGGCGAGATGCTGCTGCTGAGCGTGCAGCGGAGTGCCTGCTGCTGCTC 972

Qy 961 cgcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1020

Db 973 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1032

Qy 1021 aggcctgagatatacctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1080

Db 1033 AGCGTGAAGATATCATCCAAAAAGCTGCAAAAAATGACAAATACCTTACAGAGAGTGA 1092

Qy 1081 tatttgatctctgctgagagagcttaactccctcctcctcctcctcctcctcctcctcctc 1140

Db 1093 TATTGATCTCTGTGAGAGAGCTAAATCCCTGAGACAGAGAGAGAGAGAGAGAGAGAGAG 1152

Qy 1141 tctcagagactcagagatctgctcctcctcctcctcctcctcctcctcctcctcctcctc 1200

Db 1153 TGTTCAGAGCTCGGAAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1212

Qy 1201 cctcagagagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1260

Db 1213 CCTCAGTGGGTTACTTCT 1272

Qy 1261 tgaactgtctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1320

Db 1273 TGAAGTGTTCCT 1332

Qy 1321 tggcgaagcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1380

Db 1333 TGGCAACCTCTCGGAGATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1392

Qy 1381 tctcttcatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1440

Db 1393 TTCTCTTATTCACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1452

Qy 1441 tggggaatacttggggaatacttggggaatacttggggaatacttggggaatacttggggaata 1500

Db 1453 TGGGAAATTTGGGATCACCT 1512

Qy 1501 acccaacccttgctgaagagagctgctgctgctgctgctgctgctgctgctgctgctgctgct 1560

Db 1513 ACCCAACCTCTGTCAG 1572

Qy 1561 gcatcattgcccccttacttcttctcctcctcctcctcctcctcctcctcctcctcctcctc 1620

Db 1573 GCATCATTTGCCCCCTACTTTGTTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1632

Qy 1621 tcatggtgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1680

Db 1633 TCATGGGAGTGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1692

Qy 1681 gaaagactctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1740

Db 1693 GAATGACTCTTCCAGAAACCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1752

Qy 1741 aaaaaaacagagactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1800

Db 1753 AAAAAACAG 1812

Qy 1801 gaaaaaatatctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1860

Db 1813 GAAAAAATATCTACCTCCTTTGGTGGTAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1872

Qy 1861 atcgctgtctcccaactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1920

Db 1873 ATTGCTGTCTCCACTGAATGAG 1932

Qy 1921 atcaagaatgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1980

Db 1933 ATCAAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1992

Qy 1981 ttacaacccaacattctctgagagctcctcctcctcctcctcctcctcctcctcctcctcctcct 2040

Db 1993 TTACAAACCAACCATTTCTGAGAGAGTCTCTACTCATTAATTAATGAATGAATGAATGAAT 2052

Qy 2041 aagatgcttgaagaactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 2100

Db 2053 AAGATGCTTGAAGAACCTGTTAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2112

Qy 2101 attccaatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2135

Db 2113 ATTCCAATCATCAATTAATTAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 2147

RESULT 3

AB016257

LOCUS AB016257 2083 bp mRNA linear ROD 21-DEC-2000

DEFINITION Mus musculus mRNA for OCTNL, complete cds.

ACCESSION AB016257

VERSION AB016257.1 GI:4126604

KEYWORDS OCTNL; organic cation transporter.

SOURCE Mus musculus (Strain:C57BL 6J) adult kidney cDNA to mRNA.

ORGANISM Mus musculus

REFERENCE

AUTHORS Tamai, I., Ohashi, R., Nezu, J., Sai, Y., Kobayashi, D., Oku, A., Shimane, M., and Tsuji, A.

TITLE Molecular and Functional Characterization of Organic Cation/Carboline Transporter Family in Mice

JOURNAL J. Biol. Chem. 275 (51), 40064-40072 (2000)

PUBMED 11010964

REFERENCE 2 (bases 1 to 2083)

AUTHORS Nezu, J.

TITLE Direct Submission

JOURNAL Submitted (15-JUN-1998) Jun-ichi Nezu, Chugai Research Institute for Molecular Medicine, Inc., Gene Search Program, 153-2 Nagai, Mihari, Ibaraki 300-4101, Japan (E-mail:nezu@climmed.com, Tel:81-298-30-6211, Fax:81-298-30-6270)

FEATURES

source location/Qualifiers

1..2083

/organism="Mus musculus"

/strain="C57BL 6J"

/db\_xref="taxon:10090"

/tissue\_type="kidney"

/dev\_stage="adult"

122..1783

/gene="octnl"

/note="organic cation transporter"

/codon\_start=1

/product="OCTNL"

/protein\_10="BA3626.1"

/db\_xref="GI:4126605"

/translation="MRDDEVIAFLGEMGPORLIFLUSAIIPNGFNGSVYFLAGTPERHCLPDPVNLSSWRNHSIPLETEDGHOVQSCRRRLATIANFSAMGEPGDVDLEOLBOESCIDGWEYDKDIFLSTIYEMNLVEDMDKPLTSLFPGVLCSPFSGOLSRERKRVLEFATNAVORGSPFOVIFSNMEMPYLAIVAMGQISNVYAFILGTEILSKSRITFLGCTTFFPAITMPLDPAITIRMRMLLALTLPLGLCPVLMWFIPESRMLISORRRAEAEQIIQKAKNMSIYAPAGIDPDELQELNSLKQKVIIDLFRTRNIATITVMAVLMMLTSVGFALSLNVPNHGQVYLMCFLSGLIEPAPYFAMLLRLPRYIYAGVLFMGQVLLIIOVPEDEYNFSGIWLKFGTISAPSMXYVTAEIPLTRNNAVGIETSMASHVGSIIAPFYIGAYRRLPYILMGLSLVLIIGITLFPESGYTLPLNLEOMQKVRFRGCKSTYSVDEBSPKILITRF"

BASE COUNT 477 a 549 c 548 g 509 t

ORIGIN

Query Match 61.4%; Score 1310.6; DB 10; Length 2083;  
Best Local Similarity 85.7%; Pred. No. 2.4e-241;  
Matches 1470; Conservative 0; Mismatches 239; Indels 6; Gaps 1;

OY 137 agtggaaagcatgagggaactacagagatgacgccttccctgggagctggggccctt 196  
 DB 112 AGGTGGAAACATGGGGGACTACGACGAGGTATGCTTCTGSGGCAAGTGGGGCCTT 171  
 OY 197 ccaagcgcctcatctctctctctcagcagcagcatcatcccaatgagctcaatglat 256  
 DB 172 CCAAGGCGCTCATCTTCTTGTCTCAGGGCAGATCATCCCCAATGAGCTTCAATGTAT 231  
 OY 257 gtcagctgtctctctgagggagacccggagacagcgtgtctgagatgcccgaagcggag 316  
 DB 232 GTCAATCTGTGTCTGGGGGGGACCCCGAGCACCGTTGCTGTCTGTCACTGTGAA 291  
 OY 317 ccttagcagcgccttggcgcacaacacagtgctcgcgtcgcgtcggagcagcgcgagat 376  
 DB 292 CCTTAGGACGCTCTGGGCCAACACAGCATGCCCTTGGAGACGAAAGAGACGACACAGGT 351  
 OY 377 gccccacagctgcagcgcctacccgctcgcacacatcgccaactctcgcgtcgcgt 436  
 DB 352 GCCTCAGAGCTGGCGCCGCTACCGACTGGCCACCATGCCCAACTTCTCTCGATGGGGCT 411  
 OY 437 ggaagccggagcgagcagtgagacctggggcagctggagagagagagcgtcgtcgtgag 496  
 DB 412 GGAGCCAGGACAGGACGATGGATCTGGAGACCTGGAGGAGAGAGCTGGATGGCTG 471  
 OY 497 ggaagtcagcagcagcgtctacccgtcgcacgcgtcgtgacagtggaatcgtgtgtga 556  
 DB 472 GGAGTACGACAGGACGATCTTCTGTCCACCATGTGACAGAGTGGAACTGGGTGTGA 531  
 OY 557 ggaacaactggaagtgccccctcaacacacccgtctctctgtaagcgtgtcctcgcgtc 616  
 DB 532 GGATGACTGGAAGACACCCCTCACACACCTGTTCTTCTGTAAGCGCTCTCGCGGCTC 591  
 OY 617 ctctggtcggggagcgtgtaagcagagtttgagagagagagcgtctctcgcgaacat 676  
 DB 592 CTTCGTGTCTGGGAGCTGTGACAGAGGTTGGGAGGAAAGGCTCTTGTGACACAT 651  
 OY 677 ggccttaagagactgctcagctcctcctcagatcttctcctcagctgagagatglatc 736  
 DB 652 GGCTGTGACAGACTGATTCAGTTCGTGACAGTTTCTCAACCAACCTGGAGATGTTAC 711  
 OY 737 tgtgtatttgcacgtgggagcagtgagccagatctcacaataatgtgtgagcctcaact 796  
 DB 712 TGTGTGTTTGGCATTTGGGCGATGGGCAATCTCCAACTACGAGTGGGCTTCTCACT 771  
 OY 797 aggaacaacaaatcttgcaagtcagttcgtatataatctctcaattaggagtgagac 856  
 DB 772 AGGAACCTGAATCTGTAGCAAGTGGTTCGATCATCTTCCATTAAGAGTCTGTAC 831  
 OY 857 attcttgagctgctatagctgctgcacactgttctctactcatcagagactgagc 916  
 DB 832 ATTTTGTGAATCGGCTCATGATGTCCTCGCTGTTTGCATTAATCATCAGAGACTGGAG 891  
 OY 917 gatgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 976  
 DB 892 GATGTCGTGCTGGCCCTGACACTGCTGGCTGTCTGTCTGTCTCCCTGTGTGTGTAT 951  
 OY 977 tccctaactcccgatgctgataatccagaagaagattgagagagcttgaagatataat 1036  
 DB 952 TCCAGATATCTCCCGTGGCTGATATCCAGAGGAGATTTGAGAGGCGGAACATATAT 1011  
 OY 1037 ccaaaagctgcaaaatgaaacaacagctgtacacagcagtgatattgattctgtg- 1095  
 DB 1012 CCAGAAAGCCGCAAGATGAACAGATGCTGGCGCCAGAGGAGATTCGATCCCTGAGA 1071  
 OY 1096 -----agagagcaaatcccgtaagcagcagaaacgttcaatctcggagcgttcaagac 1150  
 DB 1072 GCTACAGGAGCTAACTCTTGAACAGCAGAAAGTCAATTAACCTGAGACTTGTAGAGAC 1131  
 OY 1151 tcggaatgctcaataatgacatattgtctgtcgtcagatgagcgtcagctgagtg 1210  
 DB 1132 TCGGAACATTTGCCACCATTAACCGTGTGATGCTGTGATGCTGTGATGCTGTGATGCTGTG 1191  
 OY 1211 ttacttgcctgtctcgtgagctcctcaattacatggagatgctcactgaactgttt 1270

DB 1192 TTAATTGCTGCTGCTCTCAATGTTCCATTTACATGAGATGCTCACTGAATGCTT 1251  
 OY 1271 cctctctgccttgattgaaatccagcttcaatcaagcctgcgtcgtcattcgagcgt 1330  
 DB 1252 CCTCTCTGGCGTGAATGAAGTTCCACTTCACTTCAACAGCTGGCGTCACTGCAACCT 1311  
 OY 1331 gcccaagcgttatcatatagctgcagtgatgttctcggagagagtggtctctctcat 1390  
 DB 1312 GCCACGGAGATATATATATAGTGGGTGCTATTTCTGGGAGAGAGTGTCCTCTTGAT 1371  
 OY 1391 tcaactgttactcgttgatattactcttcatccattcatttgcgtcgtcgtcgtcgtcgt 1450  
 DB 1372 CCAAGTGTACTCGAAGATTAATTAATTTGTCATGTCATGAGATGATGCTGGGAAAT 1431  
 OY 1451 tggagatccctcgtccttctcagctgtcgtatgtcttcaactgctgagcttaccaccc 1510  
 DB 1432 TGGGATCACCTTGCCTTCCATGTTGATGTCTTCACTGCGGAGCTGTACCAACCT 1491  
 OY 1511 ggtcaggaacatgagcgtggtggttcaatccacagcgtcccaagatgtggagcatatgc 1570  
 DB 1492 GGTACGAGACATGCTGTGGGATCACCTTCATGAGCTCTCGGGTGGGAGCATATTC 1551  
 OY 1571 cccctactgttcttactcgtcgttcttcaacagagatgctgcctacatcgtatggtag 1630  
 DB 1552 CCCTATTGTTTACTGTGGCGCTATTAACAGACTCTACCTCATCTCATGCGGAG 1611  
 OY 1631 tctgactgtctcgtatggaatcttcaaccccttcttccctggaagttggaatgact 1690  
 DB 1612 TCTGACTGTCTCTGATGGAATCAACGCTTCTTCTTCTTCTGAAAGTGTGGAGTCTCT 1671  
 OY 1691 tccagaacaccttagagcagatgcagaaatgtaatgltcaagctcgtggaaaaaacaag 1750  
 DB 1672 ACCAGAAACTTGGAGAGATGAGAAAGTGAAGAGGTTCAAGTGTGGAAAAATCAAC 1731  
 OY 1751 agactcaatgagacagaaagaaatcccaaggttctcaataactgcatctcgaataat 1810  
 DB 1732 AGTCTCAGTGAAGAGAAAGAACCCCAAGGTTCTTAATTAATGATCTTAACGAGTGT 1791  
 OY 1811 ctacccattgttgaagtgaataaacaagaaata 1845  
 DB 1792 CCAAGGCACTTGGCAAACTGAAGAGAGATGATA 1826

## RESULT 4

BC010590 LOCUS 2277 bp mRNA linear ROD 12-JUL-2001  
 DEFINITION Mus musculus, solute carrier family (organic cation transporter), member 4, clone MGC:18328 IMAGE:3670138, mRNA, complete cds.

ACCESSION BC010590.1 GI:14714870  
 VERSION  
 KEYWORDS MGC.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2277)

REFERENCE Submitted (10-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

AUTHORS Strausberg, R.  
 TITLE Direct Submission

JOURNAL  
 COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 Contact: MGC help desk  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: Iohar Hennighausen Ph.D., Robin Humphreys  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villa10n@bcm.tmc.edu.  
Villa10n, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,  
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LIML at: <http://image.liml.gov>  
Series: IRK Plate: 23 Row: 9 Column: 14  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4126604.

## FEATURES

## Source

1. 2277  
Location/Qualifiers

## CDS

```

/organism="Mus musculus"
/db_xref="LocusID:30805"
/db_xref="taxon:10090"
/clone="MGC:18328 IMAGE:3670138"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI_CGAP_Mam5"
/lab_host="DH10B"
/role="Vector: pCMV-SPORT6"
115..1776
/codon_start=1
/product="solute carrier family (organic cation
transporter), member 4"
/protein_id="AAH10590.1"
/db_xref="GI:14714871"
/translation="MRDYEYIAFLGEMPPORLIIFLLSITIPNGMSVYFLAG
TPREHCVLPDVNLSSSRNHSIPLEKDGROVPSRRYRIANTAFMSMGLEPGD
VDLEIOLDESECTDGEWEDKDELESTYTERNLVCEDEMKPLTTSLEFVVLGSEFS
GOLDRGRKRVLRATMAVDTGEFVDFSTNMFVFLAIVMGQISNVVAFLIG
TEILSKSVRIIFLGVCTEPFAIGVAVLPLEAFIRMRMLLALTLPGLECYPLMF
IPESPRMLISORFAEABOIKAKANSIYAPAGIDPELELNLSKOQKYLIDL
FRFNINATITVMAVNLISVGFALSNVPLNGHGVNLSGIVLEAPETAL
LRLPRYIAGVLEWGGVLLIIVPVDNVSIGLMLKKEGITSAPMLVYFT
ALYPLTVRMNAVGITSMASRVSIIIPYVYIGAVYRLPYILMSGIYVLIIGITLLE
PFSEFVTLPENLEDMOKVAFGRCKKSTVSVDRSEPKVLITAF"
BASE COUNT      555 a      578 c      565 g      579 t
ORIGIN

```

Query Match 61.4% Score 1310.6; DB 10; Length 2277;  
Best Local Similarity 85.7% Pred. No. 2.4e-241;  
Matches 1470; Conservative 0; Mismatches 239; Indels 6; Gaps 1;

```

137 agtgggaagcgtcggaactacgaagtgatcgcttctctggcgaagtgggacctt 196
105 AGGTGGAACAATGCGGAGCTACGAGAGGTGATCGCTTCTTG6GCGAGTGGGCCCTT 164
197 ccagcgctcattctctcgtctgaagcagcatatccccaatggttcaatggtat 256
165 CCAGGCGCTCATCTCTTCTGCTACGCGCAGCATCATCCCAATGCTTCAATGTGAT 224
257 gtcagtcgtcttcctgagcggaaccccgagacacgctgtcgaagtgcggaacgaa 316
225 GTCAAGCTGTTCTTCGCGGGGAGACCCCGAGACACGTTGCTGCTGTACACTGTGAA 284
317 cctgaagcgcctgagcgcaacaacagtgtccgctgcgctgcggaacgagcgagat 376
285 CCTGAGCGCTCTCGGCCCAACAGCATCCCTTGAGAGCAAGAGACGAGCAGCAGGT 344
377 gcccaacagctgacgcgctacgcgctcgcacacatgcccacttctcgagcgct 436
345 GCCTCAGAGCTGCGCGCTACCGACTGCGCACCATGCGCAACTTCTCTGAGATGGGCT 404
437 ggaagcgggagcgagcgtggaacctgaggaacgctggaagagagagcgtcgctgag 496
405 GGAGCGAGCAGACGAGCAGTGTGAGACACTGAGCAGAGAGAGCTGCTGATGCTG 464
497 ggaagtcagcagcagcgtctactcgtccacgctgtgacgcagatggaatcgtgtgta 556
465 GGAGTACACAAAGACATCTTCTGTCACCATGTCAGACAGATGGAATCTGATGTGA 524

```

```

QY 557 ggacaactgggaaggtgcccctcaccacccctccgttctctgttagcggtgctctcgctc 616
DB 525 GGATGACTGGGAAGACACCCCTCACCCACCTCCCTGTTTGTGTAGCGCTTCTGCGGCTC 584
QY 617 cttcgtctcgggagcagctgtcagacaggttctggcagaagaacgttctctcgcaacat 676
DB 585 CTGCTGTGCTGGGAGCGCTGTACAGACAGTGTGGCAGGAAGAAGTCTCTTTCGACACAT 644
QY 677 ggcgtgacagactggtctcagctctctcgagatlttccatcaagctggagatgttcaac 736
DB 645 GCGTGTACAGACTGTGATTCAGCTTCGCTGATGTTTCTCAACCAACTGGGAGATGTCTAC 704
QY 737 tgtgtatttgcacgtggtggacatgggacagatctccacatggtgtgagcttact 796
DB 705 TGTGTTGTTGGCCATTGTGGGCATGGCCGAGATCTCAACTTACGTGTGACCTTCACT 764
QY 797 aggaacagaatctctggcaagtcagtcgtatatacttccatctcaatctgaagtgtagac 856
DB 765 AGGAACAGAAATCCTGAGCAAGTGGTTCGATCATCTTCCACANTTACAGATCTGTAC 824
QY 857 atttttgcagtggtgctatatactgctgcacagctgttcttacttaacacagagatggcg 916
DB 825 ATTTTGTGCAATGGCTACATGGTCCGCGCTGTTTGCACTTCACTCAGACACTGGAG 884
QY 917 gatgctgctgctgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 976
DB 885 GATGCTCTGCTGCGCCCTGACACTGCTGCGCGCTGTTCTGTTCTCCCTGAGTGCTTAT 944
QY 977 tctgtaactctcccgatgctgatactccagagaagaattagaagagctgaagaatatac 1036
DB 945 TCCGAATCTCCCGGCTGATATCCAGAGAGATTTCAGAGGCGGAGACAGATCAT 1004
QY 1037 ccaaaaagcgtgcaaaaatgacaacaacagctgtgacagcagatgtatgtatctgtg- 1095
DB 1005 CCAGAAAGCCGCCAAAGATGACAGCATGCTGGCCAGCAGGAGATATTCATCTCTTAA 1064
QY 1096 ----aggaactaattccctcgaagcagcagaagaagcttcaatctgagacgtgttcagagc 1150
DB 1065 GCTACAGAGACTTAACCTCCTGTAAGCAGCAGAAAGCATATCTTGACCTGTTCAGGAC 1124
QY 1151 tctggaatatgccaataatgacatatagtcttctgctgctatgtagtgcgaactcaatggg 1210
DB 1125 TCGGAATGTCGCCACCATACCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1184
QY 1211 ttaacttgcctctctcctgagatgtctcciaattacatgagatgagcttgcgaactgtt 1270
DB 1185 TTACTTGGCTGTCTCTCTCAATTTCTCAATTTCAATTTCAATTTCAATTTCAATTTCAAT 1244
QY 1271 cctctcgtcctgattgaaattccagacttaactaagcctgctgctatctgcaagcgt 1330
DB 1245 CCTCTCTGGCGCTGATGAAATTCACACTTCACTCAACAGCCTGCTCTACTGCAACCT 1304
QY 1331 gcccaagcgttatatactgagctgcaatgctgtctgaggaagagtggtctctctat 1390
DB 1305 GCCACGGAATATATATATACCTGGGTGCTATCTGGAAGAGGTGCTCTCTCTGAT 1364
QY 1391 tcaactgtaactgtgagatattacttacttacttacttacttacttacttacttacttact 1450
DB 1365 CCAAGTGTGCTGAGGATTTATTAATTTGTCATTTGATTTGATTTGATTTGATTTGATTTG 1424
QY 1451 tgggatacactcgtcttctcactgctgatagtcttcaactgctgagcttaccacacct 1510
DB 1425 TGGGATCACCTTGCCTTCTCATGTGTATGTATGTATGTATGTATGTATGTATGTATGTAT 1484
QY 1511 ggtcagaagaacatgaggtgggtggtggtggtggtggtggtggtggtggtggtggtggtg 1570
DB 1485 GGTACGAACATGCTGTGGGATACCTTCACATGCGCTCTCGGGTGGGAGCATATTGC 1544
QY 1571 cccctacttgttactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1630
DB 1545 CCGCTATTTCGTTTACCTGGGCGCTATTAACAAGATCTCACTCACTCACTCACTCACTCACT 1604
QY 1631 tctgactgtcctgattggaattcttaacaccttttttccctgaagattgggaagtactct 1690

```

|||||  
Db 1605 TCTACTGTCCTCATTTGATCATCATCGCTTTTCCCTGAAGTTTGGAGTACTCT 1664  
Qy 1691 tcgagaaccttagaacgaatgacagaatgaaatggttcacatcgtcggaaataaag 1750  
Db 1665 ACCGAGAACTGTGACAGATGCGAAGAGTGAAGGTTTCAGATGTGGAAAAATCAAC 1724  
Qy 1751 agactcaatgagacagaaataatcccaaggttcctaactgcatctgaataaata 1810  
Db 1725 AGTCTAGTGGACAGAGAAAGAGCCCAAGSTCTAATAACTGATCTTAACGAGGTTT 1784  
Qy 1811 ctaccatttggtgaagtgaataaataaata 1845  
Db 1785 CCAAGGCACTTGGCAAACTGAAGCAATGTATA 1819

RESULT 5  
AF169831 2258 bp mRNA linear ROD 05-JUN-2000  
LOCUS AF169831  
DEFINITION Rattus norvegicus organic cation transporter OCTN1 (Octn1) mRNA,  
complete cds.  
VERSION AF169831.1 GI:5679325  
KEYWORDS Norway rat.  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 2258)  
AUTHORS Wu,X., Georgeb,R.L., Huang,W., Mangan,H., Conway,S.S.,  
Leibach,F.H. and Ganapathy,V.  
TITLE Structural and functional characteristics and tissue distribution  
pattern of rat OCTN1, an organic cation transporter, cloned from  
placenta  
JOURNAL Biochim. Biophys. Acta 1466 (1-2), 315-327 (2000)  
MEDLINE 20286310  
REFERENCE 2 (bases 1 to 2258)  
AUTHORS Wu,X., Wang,H., Leibach,F.H. and Ganapathy,V.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-1999) Department of Biochemistry and Molecular  
Biology, Medical College of Georgia, 1120 15th Street, Augusta, GA  
30912-2100, USA

FEATURES  
source  
1..2258  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
1..2258  
/gene="Octn1"  
72..1733  
/gene="Octn1"  
/codon\_start=1  
/product="organic cation transporter OCTN1"  
/protein\_id="AAD46922.1"  
/db\_xref="GI:5679326"  
/translation="MRDYDEVIAFLPDMGPRORLIFLLSAILIPNGFNMSVPLAG  
TPRHRIYHYNYSSARNSPILETGDRQVRSRRRIATNPSLGEPLD  
VDLELDQESCDGWPIKDFLSTTVENMLVGEDMKVPLTSLFVGLGVS  
GQLSDRGKRVLFATMAVOGEFVOLFSTNMEFVLAIVMGQISNVVAFLIG  
TELISRVRLISTLGCTFFAIGYMLPLFAFIRRMRLALATPLGFCVPLWF  
IPSPRLMISQRFEEADQIIOKAKKNGIMAPVIFDPLELQSLNKOQYIIDL  
FKRNATITVMSVLMMLTSGVFPALSLNPNHGDVYLMCFGLIEVPAFTML  
LRTLEPRYIIVGLFVGSGVLLVQVAPEDYINFSIGVLMGFGVTSFMSLYFT  
ALYPLVYNNMAVGITSMASRGSITAYFVYLCAYNRLLPYILMGSITVLIQITLFP  
FPESRGYITPEMLQOMOKYRFRCKKSTVSMDBENPKVLTAF"

BASE COUNT 558 a 565 c 566 g 569 t  
ORIGIN

Query Match 60.6%; Score 1293; DB 10; Length 2258;  
Best Local Similarity 85.1%; Pred. No. 5.7e-238;  
Matches 1459; Conservative 0; Mismatches 250; Indels 6; Gaps 1;

Qy 130 gagcgcagctgggaagcatcggaactacagcaggtgactccttcctgctgagctgag 189  
Db 55 GAGGAGAGCGCGGAAGCATGAGGAGACTATGACGAGGTGATAGCCTTCCTGGGCGATTGG 114  
Qy 190 ggccttcagagcgcctcatctctctctgctgacgagcagcatatcccaatgcttca 249  
Db 115 GGCCTTCACAGCGCCTTATCTTCTCTGCTCAGCCACAGCATCATCCCAATGGCTTCA 174  
Qy 250 atgatatgctcagctgctctcctgagcggagcccgagacacgcctgctgagctgcgagc 309  
Db 175 ATGGATATGATGATGCTGCTGCTGCGGGGACCCGAGAGCATGTGGCTTGGCTTCACA 234  
Qy 310 ccggaacctgagcagcgcctgagcgaacaagtgctccgcgtgcgctgcgagcggcc 369  
Db 235 CCGTGAACCTTGAGCAGCGCGCTGGCGCAACACAGCATCCCTTGGAAAGCAAGCGAGAC 294  
Qy 370 gcaagctgccccacagctgcagccgcctcagcctcgcacatcgcacatctcggagc 429  
Db 295 GACAGGTGCTCTCAAAAGCTGCCCGCGCTACCGACTGGCCACATCGCCAACTTCTGCGGC 354  
Qy 430 tcggcctggaagccgggagcgtggaactgagcagctggaagcagagctgcctgg 489  
Db 355 TGGGCTTGAGCCCGGACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 414  
Qy 490 atgagctggaagctgagcagcagcctgacacgctgacacgctgagcagcagctgag 549  
Db 415 ATGGCTGGAGTACAGCAGGAGCTCTTCTCTGTCACATCTGTCAGAGTGGAAATCTGG 474  
Qy 550 tgttgaggaacactggaaggtgcccccaacacccctcctgctgctgagcagctgctcc 609  
Db 475 TGTGTGAGAGATGACTGGAGACACCCCTCACACCTCCCTGTTCTTCTGAGGCGTTCTCT 534  
Qy 610 tggctcctctgctgctcgggagcgtgacagcaggttgtagaggaagagcgtctctg 669  
Db 535 GTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 594  
Qy 670 caacatgctgctacagactgctcagctcagctcctcagatattctccatcagctgagga 729  
Db 595 CCACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 654  
Qy 730 tgttactgcttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 789  
Db 655 TCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 714  
Qy 790 tcaactagagacagaatctctgcaagctcagctcgtatlatatctcctaactaattagag 849  
Db 715 TCATATCGGGAGCGAATCTGCTGAGCAAGTCGCTGATCTCTCTGCTGCTGCTGCTGCTG 774  
Qy 850 tctgcatcttttctgagctgctgctgctgctgctgctgctgctgctgctgctgctg 909  
Db 775 TCTGATCATTTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 834  
Qy 910 actgagcagatcgtcgtgctgctgctgctgctgctgctgctgctgctgctgctgctg 969  
Db 835 ACTGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 894  
Qy 970 ggttcatcttcgaatctcccgatgctgctgctgctgctgctgctgctgctgctgctg 1029  
Db 895 GGTATTTATTCAGAGCTTCCTCCGCTGCTGATATCCAGAGAGATTTTGAAGGAGCAAC 954  
Qy 1030 atcatcacaagaacgctgcaaaaatgaacaacagctgctgcaagcagagatattgatt 1089  
Db 955 ACATCATCCAGAAAGCCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1014  
Qy 1090 ctgtg-----aggactaaatccctcctgagcagcagaagcttcatctcgtgacctgt 1143  
Db 1015 CTTGAGACTACAGAGGATTAACCTCTTGAAGCAGCAAGAAAGCTTCATCTGAGCCTGT 1074  
Qy 1144 tcagactcgaataatgccataatgacattatgcttctgctgctgctgctgctgctgctg 1203  
Db 1075 TCAAGACTCGAAGACTTCCACATATACAGTGAATGCTGTGATGATGCTGTGATGATGCT 1134  
Qy 1204 cagtggttacttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1263

[illegible][illegible]





OY	400	ggctcgccacacatgccaactcttcaggcgtctggagctggagccggagcgccgaagtggacc	459
Db	475	GGCTCGCCACCATGCGCAACTCTTCGGCGCTCGGGCTGGAGCCGGGGCCGCAGAGTGGACC	534
OY	460	tggggcagcttggagcaagaagctgccttgtagtggcttggagattcaagccaagcgtctacc	519
Db	535	TGGGGCAGCTTGAGCACAGAGAGCTGTCTGGATGGCTGGGGATTCCACTTGAGAGAGTCTTACC	594
OY	520	tgtaaacacctgtagccagatggaatactctgtgtgtgaggaacaacttggaaagtgtgccctta	579
Db	595	TGTCACACATTTGTACCGGAGAGAGACTGGTGTGTAGAGAACACTGGAAAGCCCCACTCA	654
OY	580	cccaactecctgtctctctgtagagcgctgcctccctggcctctcgtgtcccgaggcagctgtcaa	639
Db	655	CAACTCTCTTCTTCTTGGTGGGTGTGCTGTGGGCTCTTTCATTTGAGGCCACTGCTCAG	714
OY	640	aagaatttggcaggaagaacgctctctctccaacccatgctgtcatgagactggcttcagtt	699
Db	715	ACAAGTTTGGCGGGAAGAAGTGTCTGTCTTGACACATGGGAGCAGACAGAGCTTAGCT	774
OY	700	tccctgcagaatctctccatcagctctggagaatgttcaacttgttatattgtcatctgtggcca	759
Db	775	TCCCTGCAATATTCTTCGAAGAATTTTGAATGATTGTGCTGTGCTGTGTGCTTGTAGGCA	834
OY	760	tgggccaagatcccacaatatgtgttagccttcacactagaagaacaagaattctctgcaagt	819
Db	835	TGGGCGAAGATGTCACACTATGTGGCAGACANTTGTCCGGGGAGAAAATTTCTTGGCAAGT	894
OY	820	cagttctgtatataattctctcacatagaagtgtgcacaatttcttgcagttgtgctatagc	879
Db	895	CAGTTCGATAATATATCTCTACGTTAGAGAGATGTCAATATTTATMGCAATTTGGCTACATGG	954
OY	880	tgctgcacactgttgccttaacttcaatcagagactggcggagatgcctgccttggcgccttacgg	939
Db	955	TGCTGCGCACTGTTGCTTACTTCAATCCGAGACTGGCGAGATGCTGTGTHGGGCGCTGACGA	1014
OY	940	tgcgcggagatgctgtgtgtgtcccgctgtgtgtgtgttcaatctcttgatattccccgatgtcta	999
Db	1015	TGCGGGGGTCTGTGGTGCGCACTCTGTGTGTTCATCCTGTGATCCCGCCGATGGCTCA	1074
OY	1000	tatcccaagaagaattttaagaaggctgaaagatatcatcacaanaaagctgcacaaaatlaaca	1059
Db	1075	TCTCTCAGGGGCGATTTTGAAGAGCAAGGTGATCATCCGCAAGGCTGCCAAGGCCAATGG	1134
OY	1060	acaagcgtgtccagcagatgatttattgattctgtgagag-----ggagctaaatcccctba	1113
Db	1135	GGATTGTTGTCCCTTCCACTATCTTTTAACCCGAGTGAAGTTTACAAGACTTAAGTTCCAABA	1194
OY	1114	agcagcagaagaacttcaatcttgcagccgttccagagactcgnaatatgtccataatgacca	1173
Db	1195	AGCAACACTCCACACAACATTCGTGATCTGCTGCAACCTGGAATATCCGATGGTACACA	1254
OY	1174	ttaagtcttgcctatgtagtgcagccctcagtgaggttacttgcctcgtctctctctgtagt	1233
Db	1255	TCAATGTCATATATGCTGTGATGATMCATATCAGTGGGCTATTTTGGGCTTTGCCCTTGATA	1314
OY	1234	ctctcaaatltaacatgtagatgctcactcagaaactgtttctcctctcgtccttgattgaatco	1293
Db	1315	CTCTCAACTTCATCGGGGACATCTTTTGTGAACCTGCTCTTTCAGGAGATGGTTGAAGTCC	1374
OY	1294	cagcttacaatcaagccgtgcgtgtatgttgaaagcctgcggcagagcgltatatatagctg	1353
Db	1375	CAGCAATATGTTTGCTGGCTGGCTGCTGTGTGAATATTTGGCCCGGCGCTATTCATATGGCA	1434
OY	1354	cagctatgtctctgggaggaagatgtgctctctctcaatcaactgtgtaacctgtgattatt	1413
Db	1435	CTGGCCCTTCTTCCTGGGTGGGAGTGTCTCTTCTTTCATGACGAGCTGTGAACCCCGCAGACTTGT	1494
OY	1414	actcttlatcatcgtgtctgtcatgtctggaaaaaatttggaaacoactcgtccttctca	1473
Db	1495	ATTATTTTGGCTACAGTCTCTGTATATGTGGCAAGTTTGGAGTTCACAGCGCTCTTTTCCA	1554
OY	1474	tgtctgtatgtcttcaactgtctgtagctctaccacaacctgtctcaggaacaatgtgcgtgtggg	1533

Db	1555	TGCTCTAGTGTACACAGCCGAGCTGTATCTCCACACTGTGTGAGAAACATGGGTGGGAG	1614
OY	1534	tcacatccaagcctccagaagtgtggcaatcatctgtccccctacttctgttacctcggtg	1593
Db	1615	TCACCTCCACAGCATCCCGCTGGGAGACATCTGTCTCCCTCACTTCCTTACCTTGGTG	1674
OY	1594	cttaacaagaatgtcgtccctacatcgtcatgtgtgtatgtcgtacgtgcctgattgaaatc	1653
Db	1675	CTTACGACACGGTCTCTCCTTACATTCTCATGTGGAAAGTGTGACACCATCTGCACGCCATCC	1734
OY	1654	tcacccttttttcccttgaaagtlttgggaatgaactcttccagaacacttaagcagatgac	1713
Db	1735	TCACCTGTTTTCTCCGAGAAGCTTGGTATCCCCTCACTCCAGACACACTTTCAGATTC	1794
OY	1714	agaaagtgaatgtgtcatcgtctgggaaaaaacaagagactcaatctgagacagaagaa	1773
Db	1795	TAAGATCAAAGAAATGAACACGAAAAACCTCAAGTGCACACAAAGATGTTAAAGANG	1854
OY	1774	atc 1776	
Db	1855	GTC 1857	

RESULT	8
BC012325	
LOCUS	2811 bp mRNA linear PRI 10-AUG-2001
DEFINITION	Homo sapiens, similar to solute carrier family 22 (organic cation transporter), member 5, clone MGC:20139 IMAGE:4548603, mRNA, complete cds.
ACCESSION	BC012325
VERSION	BC012325.1 GI:15147377
KEYWORDS	MGC.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
REFERENCE	1 (bases 1 to 2811)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (09-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT	Contact: MGC help desk Email: gcbp@remail.nih.gov Tissue Procurement: DCTD/DTP/Gazdar cDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@gsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Keta Kusche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lilja Pirabhu, Parvaneh Seedei, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Speede, Jeff Stolt, Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.  Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: http://image.llnl.gov Series: IRAL Plate: 29 Row: A Column: 13 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3273740.  FEATURES location/Qualifiers 1..2811 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="MGC:20139 IMAGE:4548603" /tissue_type="lung, large cell carcinoma"



CDs

```

/cdone_1lb="NIH_MGC_18"
/lab_host="DH10B-R"
/notes="Vector: POTB"
172..1845
/codon_start=1
/product="Similar to solute carrier family 22 (organic
cation transporter), member 5"
/protein_id="AAH12325.1"
/db_xref="GI:15147378"
/translating="MRDYDEVIAFLGEMEPFORLIFPLISASTIPNGFTGLSSVELIA
TPBHRKRVDPANILSSAMRNHTVPLRLRGRREVPHSCRRRLATIANFSALGIEPRD
VDLGLEQESCPDGMERSODVYLTITVENILCEDMDNVLISLEFVGLGFSIS
GOLSDRGKRVLEFVMQOTGFLQIFSKNFEMFVLEVLVGMQISNYVAFLG
TEILGKSVRIIFSTIGYICFARFVAVWVLPFAFIFRDMRLVALMPGVLCVALLMF
IPSPRRLISQGRFEREAVIIRKAANKNVYSTFPMSELODSSKROOSHNIIDL
LRTNIMRVMTIMSIMMTISVGFISLIDTPMLHEDIVNCELSMVEPRATVLAHL
LIQYLPFRYMATALFLGGSVLLFMQIVPDLVLAIVLWVKFVTAFSMYYIT
AELYPVVRNMGVSVSSTASRLGISLSPYVIVLAVDRFLPYILMGLSLILTAIILFL
LPSEFGPLPDTIDOMLRVGMKRRKPSHTBMLKDGQERFTLKSTAF"
BASE COUNT      632 a      743 c      726 g      710 t
ORIGIN
Query Match      49.8%; Score 1062.6; DB 9; Length 2811;
Best Local Similarity 77.4%; Pred. No. 8.9e-194;
Matches 1303; Conservative 0; Mismatches 374; Indels 6; Gaps 1;
QY 100 cagctgcaaccgtagttgcaagtttcggagcgagtgaggaaacatgcggactacg 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 GCGGTCCCCGACCCCGCGCGCGCTGTGGGCTTGAGGGCGGCACTGCGGAGTACG 184
QY 160 acgagtgatgacctctctggggagtgaggcccttcacagcctatctctccgc 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 ACGAGGTACCGCTTCCTGGGCGAGTGGGGCCCTTCACAGCGCTCATCTTCTCTGC 244
QY 220 tcagcgcaagcatatccccaatgagcttcaatgtaatgtaatgtaatgtaatgta 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 TCAGGGCGAGCATCATCCCAATGAGCTTCACGGCGCTGCTCCGTGTTCTGATAGCA 304
QY 280 ccccgagcaacgcgtgtgtgagtgccgagcccgaaacctgaagcagcgcttggcgaa 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 CCCCGAGACCGCTGCGGGGTGCGGAGCGCGCAACCTTAGAGCGCCCGGGCGCAAC 364
QY 340 acaagtgcccgctgagctgcggagcgagtgagtgagtgagtgagtgagtgagtgag 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 AACTGTCCACTGCGGCTGCGGAGCGCGGAGGTCGCCACAGCTGCGCGCTTACC 424
QY 400 ggcgcgcaacatcgcaactctcgagctcgagcttggagcgagcgagcgagcgagcgag 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 GGCTGCGCACATCGCAACTCTCGGCGCTTGAGGCGGAGCGGGGCGGAGCGAGCAC 484
QY 460 tgggagcagctgagcagagagagctgcgtgagtgagtgagtgagtgagtgagtgag 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 TGGGGCAGCTGGAGGAGGAGCTTCGGAGTGGCTGGAGTTCAGTAGGACGTCTACC 544
QY 520 tgcacacgctgagcagagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 TGTCCACCTTTGTGACCGAGTGAACCTGTGTGAGGAGCAGCTGGAAGGCCCACTCA 604
QY 580 ccaactccctgtctctgtaagcgctgctcgcgtctcctctgctgcggagcagctgcaag 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 CATCTGCTTGTTCGTTCGTGGGTGTGTGGCTCCCTCATTTACAGGCGAGCTGTAG 664
QY 640 acaagtttggcaggaagacgctctcttcgcaacatgagctgtaagactggtctcaact 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 ACAGGTTTGGCGGAGGAATGCTGTTCTGACCATGGGCAATGAGCAAGGCTTTCACCT 724
QY 700 tccctgagaatttccacacagctggagaatgtaactgtaactgtaactgtaactgtaact 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 TCTGTGAGATCTTCTGAGAGAAATTTGAGATTTTGTGCTGTGCTTGTGTAAGGCA 784
QY 760 tgggccaagatctccaactgtgtagcttatactaggaacaaactcttgcaagt 819
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 785 TGGGCCAGATCTCCAACTATGTGGCAGCATTTGTCTCGGGGACAGAAATTTGCGCAAGT 844
QY 820 cagttgtaattatctctacataggagtgacacatttttcagttgtaattatgc 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 845 CAATTGCTAATATTTCTCTACGTTAGAGTGGATTTATTTATGATTTGCTTCATGG 904
QY 880 tggcgcacagtttggcttactcaatcaagagcttggcgagatgctgctgctgctgag 939
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 TGTGTGACAGTTTGTCTTACTTCTCATCCGAGACGTGGGAGTGTCTGTGGTGGCGGTACGA 964
QY 940 tgcggagagtgctgtgtctccgcgtgtgtgtgttaattccgtaatcccgagtgctga 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 965 TGCCTGGGGGCTGATTCGTGGCACTCTGTGTGTATCCCTGATGCCCATGGCTCA 1024
QY 1000 taccccaagaaatttagaagagcgtgaagatatcatccaagaaagctgcaaaatgaaca 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1025 TCTCTAGGGACGATTTGAGAGAGGAGAGTGTATCTCCGCAAGGCTGCCAAAGCCAAAG 1084
QY 1060 acacaagctgaccagagatgatatgtgattctgtgag-----gagctaaatccctga 1113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1085 GGAATGTGTGCTTCACATATCTTGACCCGAGTGATGATACAGACTAAGTCCAAAGA 1144
QY 1114 agcagaagaagcttcaatctgagacctgttcaagactggaatatgtccataatgacca 1173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1145 AGCAGCAGTCCCAACACATTCGTGATCTGTTGGAACCTGGATATCCGATGTGTACCA 1204
QY 1174 ttatgtcttgctgtatgatatgtgacctcagtgagttgtaactgtctgtctcgtgagtg 1233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1205 TCATGTCCATTAATGCTGTGATGATACCATATCAATGGGCTAATTTGGGCTTTGCTTGANA 1264
QY 1234 ctccaaattacatgagatgagctgaactgaactgtctctctctgtccttgatgaaatc 1293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1265 CTCTCACTTGCATGATGGGGAATCTTGTGTAATGCTTCCATTGAGGATGTTGAAGTCC 1324
QY 1294 cagcttaccatlaacagctgtgctgtatcttcgaaagcgtgcgcaagtggtatatagctg 1353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1325 CAGCATATGTTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1384
QY 1354 cagtaactgtctgagagagagtggtgtctctctctcaactgaactgtgactgtgattatc 1413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1385 CTGCCCTTCTCTGAGGTGGAGAGTGTCTTCTTCTTCAAGCTGTGATCCCCAGACTTGT 1444
QY 1414 actctatcatcatgtgtctgtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1445 ATTAATTTGGCTACAGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1504
QY 1474 tgcgtatgtcttcaactgctgagactctcaaccacccgtgtcagaagaatgagtgagtgag 1533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1505 TGGTCTACAGTGTACACAGCGAGCTGTATCCACAGTGTGAGAAACATGGGTGTGGAG 1564
QY 1534 tcacatcaagagcctcagagtgagcagagatcatgtgccctcaacttgttttaactcgtgtg 1593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1565 TCACCTCACACAGCATCCCGCTGGGAGCATCCGTGTCTCTTACTTGTGTTTACTTGTGTT 1624
QY 1594 cttaacaagaatgctgcccacacatgcatgagtgagtgagtgagtgagtgagtgagtgagtgag 1653
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1625 CCTACGACCGCTTCTCTCCCTACATTTCTCATGGGAATGTGACCATCTCAGACCATCC 1684
QY 1654 tcacccctttttccctggaagtggttggaatgactcttcagaagaacttgaagcagatgtg 1713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1685 TCACCTTGTTCCTCCAGAGAGCTTCGTACCCCTACCTCCAGACCATCTTGTGACCATGTC 1744
QY 1714 agaaagtgaatggttcagatctgaggaagaaagaaagaaagaaagaaagaaagaaagaa 1773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1745 TTAAGATCAAGAAATGAACACAGAAAACTCCAAATCACACAGATGTTAAAGATG 1804
QY 1774 atc 1776
    ||
Db 1805 GTC 1807
    ||

```

RESULT 9  
AF111425



1504 caaccctggtgaagaacatggtggtggtcacatccacggcctccagagtggtgacga 1563  
1463 CCACGTGTGTCAGAAACATGGGTGTGGGGTACCTCCACGACATCCCGCTTGGCAGCA 1522  
1564 tcaatgcccctactgtttactcctcggtgttacaacagaatgctccctacatcgta 1623  
1523 TCCTGTCTCCCTACTTCTTTACTAGGTGCTGATGATCGCTTCCCTGCTTATATCTCA 1582  
1624 tgggtagtcgtactgtcctgtatggaaccttcacccctttttccctgaaagttggaa 1683  
1583 TGGGAAGTCTGACATCTGACATGACATCCCTGCTTCTCTCCCTGAGAGCTTTGGTG 1642  
1684 tgaccttcacgaacaccttagacagatgacgaagaatggtgtgaatcgtcgtgga 1743  
1643 TCCCTCTCCAGATACCAATTAACGATGCTAAGGGTCAAAAGATATAAACAATGGCACA 1702  
1744 aaacaagactcaatgtagacagaaga 1771  
1703 TCCAAAGCCAGACAAAGATGCAAAAAGA 1730

RESULT 10  
AB015800 1888 bp mRNA linear ROD 19-JUN-1999  
LOCUS AB015800  
DEFINITION Mus musculus mRNA for OCTN2, complete cds.  
ACCESSION AB015800  
VERSION AB015800.1 GI:4126482  
KEYWORDS octn2; OCTN2.  
SOURCE Mus musculus (strain:C57BL 6J) adult kidney cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.  
1 (sties)  
Nezu,T., Tamai,T., Oku,A., Ohashi,R., Yabuuchi,H., Hashimoto,N.,  
Nikaido,H., Sai,Y., Koizumi,A., Shoji,Y., Takada,G., Matsushita,T.,  
Shimane,M. and Tsuji,A.  
Primary systemic carnitine deficiency is caused by mutations in a  
gene encoding sodium ion-dependent carnitine transporter  
Nat. Genet. 21 (1), 91-94 (1999)  
99113835  
2 (bases 1 to 1888)  
Nezu,T.  
Direct Submission  
Submitted (24-JUN-1998) Jun-ichi Nezu, Chugai Research Institute  
for Molecular Medicine, Inc., Gene Search Program, 153-2 Nagai,  
Nihari, Ibaraki 300-4101, Japan (E-mail:nezu@climmed.com,  
Tel:81-298-30-6211, Fax:81-298-30-6270)  
location/Qualifiers  
1. 1888  
/organism="Mus musculus"  
/strain="C57BL 6J"  
/db\_xref="taxon:10090"  
/tissue\_type="kidney"  
/dev\_stage="adult"  
60. 1733  
/gene="octn2"  
/gene="octn2"  
/codon\_start=1  
/product="OCTN2"  
/protein\_id="BAA36590.1"  
/db\_xref="GI:4126483"  
/translation="NRDYDTAFIEMGCPFORILFLLSLIIPNGNSIYFIAG  
/PERHCLVPTVNSSANRNHSTIPLETGDKQVPOKCRKRIATIANSELGERD  
VDLEQDESCLDGEWYDKDFLSTIVEMDLVCKDDKAPLTSLEFVGVMKSFIS  
GQSDRFGRKNVFLTMGMQGFSEFLQVSVNFEMFTVLFVLVGGQISNVVAFLVIG  
TEILSKIRIIPATIGVCIFYAFGPMVPLFAFYTRDMRLMLALATVPGVIGALMWF  
IPESPRMLISGRKEAEVYIIRKAKINGIYAPSTIPDSELDONSTKPOLHIYDL  
ITRNIRVTITMSIILKLTISVGRGLSLDIPDHGDIYVNCFLIAAEVPAVYIAML  
LTQYLPRTSTSAALPLGGSVLLFMQVLPSELPIYSTRALVWVKRGITISATMYVT  
AEIYPTVVRNMGVSVSTASRLGISLSPFYVLGAYDRFLPIILMGSILTLLATITLF

BASE COUNT 392 a 509 c 486 g 501 t  
ORIGIN  
Query Match 46.6%; Score 94; DB 10; Length 1888;  
Best Local Similarity 75.7%; Pred. No. 1.2e-180;  
Matches 1247; Conservative 0; Mismatches 395; Indels 6; Gaps 1;  
PFSPGVLPPTIDQMLRVKIGIKOMQIOSQTRMOKDEESEPVLKSTAF"  
130 gacgcgaatggaagaacatgctgaggaactaagaagaatgctgcttcctcggcagagtg 189  
43 GGGAGGCTTGAGAGCGGCGATGCGGAGCTACGACAGAGGTGACCCGCTTCTGAGCGAGTGG 102  
190 ggccttcacgagcccatctcttcctcgtcaagcgccagcatcaccacaaatggtctta 249  
103 GGGCTTCCAGCGCCCTCATCTTCTTCTGCTACGCGCCAGCATCATCCCATAGGCTTCA 162  
250 atggtatgcatcgtgctgctcgtgagggagaccccgagagacgctgctgagtgcgagc 309  
163 ATGATATGTCATCGTGTCTGCTGGCGGGACCCCGAGACACCTGCTTGTGCTTCACA 222  
310 ccggaacacctgaagcagcctgagcgaacaaagctgctcgctgagcgtgcggagagcgcc 369  
223 CCGTGAACCTGACGACGCGCTGGCGCACCAAGATATCCGTTGAGAGACGAGACGAC 282  
370 gcgaggtgcccacgaagctgagcagcctacccgctcgcacacatcgcacacttcctcgagc 429  
283 GACAGGTGCTCAGAAATGCGCGCGCTACGAGCTGGCGACATCGCAACTTCTCGAGC 342  
430 tgggtcgtgaagcggcgagcgtgagccttggggaagctgtgagcagaagaagctgcctg 489  
343 TAGGGCTGAGACGGGCGGGAGCTGAGACCTGAGACAGCTGGAGAGAGCTGCTGG 402  
490 atgctcgtgaggtcagcagaagcgtctacgtgcaacgctgcagcagatggaatcgtg 549  
403 ATGCTGGGATACGACAAAGAGAGCTCTCTGTCACCATGAGACAGTGGAGCTGG 462  
550 tgtctgaggaacactggaagtgccctcacacacactccctgtctctctgtagtcgtcc 609  
463 TGTATAGATGATGATGAGAAAGCCCACTACACACCTCTGTTTGTGGTGGTGTGTA 522  
610 tgggtcctcctgctgctcggcgaagctgtcagaagaagtttggcagaagaagctccttg 669  
523 TGGGCTCTTCATTTTCAGACAGCTTCACAGAGTGTGGTCCCAAAATGCTCTGTTT 562  
670 caaccatggtcgtacagaactggtcagcttcctcgtcagatcttcctcacaagctggaga 729  
583 TGACCATGGGCATGACAGATGCTTTCAGCTTCTGCGAGGCTTCTGTGTAACCTTGACA 642  
730 tgtcactgtgttatctgtcactcgtgagcatggtggccagatctccaactatgtgtagcct 789  
643 TGTTTACAGTGTCTTTTGTCTCTGTTGGCANTGGGTGAGATCTCCAACTAGGTGAGCAT 702  
790 tcaataagagaacagaatctcttgcaagatcgtgatatataatcttcaacattggag 849  
703 TTGTCTCGGAAACAGAAATTTCTTCCAAAGTCAATTAATTAATTAATTAATTAATTA 762  
850 tgtcacacttttggcagttggtgctatagtctgctgcacacgtttgttacttaacagag 909  
763 TTTCATATTTTATATGCGTTTGGCTTCAATGAGTGTGCTGCACTTTGCAATTCATCGAG 822  
910 actgacgagatgctgctgctgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagc 969  
823 ACTGAGAGATGCTGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 882  
970 ggtcactcctgaatctcccgagatgctgatatccagagagaagttaagagagagag 1029  
883 GGTTCATCCCTGAGTCCCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 942  
1030 atatcatcaaaaagctcgaaaaatgaaacaaacagcgtgtacacagagtgatattgatt 1089  
943 TGATCATCGCAAAAGCTGCCAAATTCATAGGAGATGTTTGGACCTTCCACATATCTTGATC 1002



D6 636 TGACCAAGGCGATGACAGACTGGCTTACGCTTCCATGCAAGGCTCTCTCTGTAAGCTTCGAGA 695

QY 730 Tgttactgtgttattgttcacgttgagcattggccagatctccaaactatgtgttagcct 789

D6 696 TgTTACAGAGGCTTTTGTCTCTGTTCATGGGTGCGATGAGATCTCCAACTAGCTGGCGAT 755

QY 790 tcataatgaagaacaaatctcttggaagtcagttcgtatatactctacatagag 849

D6 756 TTGTCCTGGGACAGAAATTTCTTCCAAAGTCAATTCGAATTAATATGCGCACCTTAGAG 815

QY 850 tctgcaacatttttgagagttgctatagtctgtccacgtttgttactatcaag 909

D6 816 TTTCATATTTTATGCTTTGGCTTCAATGAGTGGCTGACACTGTTGGCACTTATATCAAG 875

QY 910 actgagcagatgtgtctgtgagcaggtgcgcgagtgctgtgtgtccgcgtgtgt 969

D6 876 ACTGAGAGATGCTGCTGCTGGCGCTCAGTGGCAGGGGTCATGTGGGGCTCTCGT 935

QY 970 ggttcaattcctgaatcctcccgatgtgctgatatccagagaagattagaagagctgaag 1029

D6 936 GGTTCATCCCTGAGTCCCGACAGATGCTCATCTCTCAAGGCCGAATTAAGAGGAGAGG 995

QY 1030 atatcccaaaaagctgcgaaaaatgaacaacacagctgtaccagcagctgatatgttgc 1089

D6 996 TGATCATCCGCAAGGCTGCCAAATCAATGAGATGTTGGCACCCTTCACATGCTTGATC 1055

QY 1090 ctgtgag-----gagctaactcccttgagcagcagagaagctttcattctgagctgt 1143

D6 1056 CAAGTGAATGACAGACTTAAATTCACGAAGCTCAGTGGCACACACATTATGATCTGA 1115

QY 1144 tcagagagcgaatattgacataatgacataatgcttctgtctgtctatgtgagctgacct 1203

D6 1116 TCCGAACACGGAATATCAGGGTATCACCATCATGTCATATCTATATCTGTGCTGACCATTAT 1175

QY 1204 cagtgaggttactgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 1263

D6 1176 CAGTGGGCTTATTTGACATATCTCTGTGACATCTCTTAACTTGAAGGGGACATCTATGTA 1235

QY 1264 actgttccctctctgtctgtctgtgtgaatctcaagcttaccattacacagctgtctgtctgt 1323

D6 1236 ACTGCTTCCTACAGGCGGCTGTGAAGTCCGACGCTATGTCGCGCTGTGTGTGCTG 1295

QY 1324 gaacgtgcacagcggtatatacagctgtgacgtgtctgtgtgtgtgtgtgtgtgtgtgt 1383

D6 1296 AGTACTGCCCCCGCGGATATCTATCTCGGCTGCCCTTTCCAGGGGCGAGGTCTCTC 1355

QY 1384 tcttcatcgaactgtgtaaccgtgtgattatattcttctatccattgtctgtgtcatgtgtg 1443

D6 1356 TCTTCATGACGCTGTGCTTCAAGAAATTTGTTTAACTTGTCTCCTGCTGATGCTGTG 1415

QY 1444 gaaatttgagatcaactctgtctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 1503

D6 1416 GGAAGTTTGAATACCTCTGCGCTACCTCATGCTATGCTGTACACAGCTGAGCTGTGACC 1475

QY 1504 caaccctgttgaagaacatgagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1563

D6 1476 CCACTGTGTCAGAAACATGGGTGTGGGGTGCAGCTCCACAGCATCCCGCTTGGCGACA 1535

QY 1564 tcaatgtccctctacttctgttactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1623

D6 1536 TCTCTGTCTCTTCTTGT 1595

QY 1624 tgggttagctgtgactgtctgtatgtgaatcttcaaccccttttccctgtgaagtgtgtgaa 1683

D6 1596 TGGGAATGTGACATCTGTGACAGCTATCTCTACACTTGTCTTCCGGAAGCTTTGTG 1655

QY 1684 tgaactcttccagaacacttgaagcagatgcaagaagtgaatgtgttcaaga tctgtgtgaa 1743

D6 1656 TCCCTCTCCCGATACCATGTGACAGATGCTAAGGGTCAAGAAGAACTGAAGTGGCAAA 1715

QY 1744 aaacaagaactcaatgt 1771

D6 1716 TCCAAAGCCGACAGAAATGCAAAAAGA 1743

RESULT 12  
RNAJ1933 3007 bp mRNA linear ROD 30-MAR-1998  
LOCUS Rattus norvegicus mRNA for putative integral membrane transport protein (USTR2r).  
DEFINITION Rattus norvegicus mRNA for putative integral membrane transport protein (USTR2r).  
ACCESSION AJ001933  
VERSION AJ001933.1 GI:3004486  
KEYWORDS integral membrane transport protein; USTR2r gene.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 3007)  
AUTHORS Gruendemann, P.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-1997) D. Gruendemann, University of Heidelberg, Department of Pharmacology, Im Neuenheimer Feld 366, 69120 Heidelberg, FRG  
REFERENCE 2 (bases 1 to 3007)  
AUTHORS Schomig, E., Spitzemberger, F., Engelhardt, M., Martel, F., Ording, N. and Gruendemann, D.  
TITLE Molecular cloning and characterization of two novel transport proteins from rat kidney  
JOURNAL FEBS Lett. 425 (1), 79-86 (1998)  
MEDLINE 98200080  
FEATURES  
source location/qualifiers  
1..3007 /organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/tissue.type="kidney"  
87..1760 /gene="USTR2r"  
/gene="USTR2r"  
/codon.start=1 /product="putative integral membrane transport protein"  
/protein.id="CA05106.1"  
/db\_xref="GI:3004487"  
/db\_xref="SPTREMBL:O70594"  
/translation="MRDYEVTAFLEGMPORLIFELLSASIIPIGNGMSIVFLAGTPENRCLVPTVNLSAARNHSIPLETGKQVPOSRORYPLATIANFSALGLEPHDVLDELQENCLDGEYKDYSLITVEMDLCKDKMKALITSLFPVGIAMGSPFSGOLSDRGRKRVNVLFTMGOMQGESFLOFSYNEPMPTVLYVMQGISVYAAFVGTSLKSKIRITFIATLGCITFAFEPNPLPFAITRDMRMLLATVPGVLCALMWFIPESRMLISQRYKKEAVIIRKAKENGIVAPSTIFDPSLQDLNKKPOSHHITDLVTRNIRITITMSIILMTISGVYFGLDPNPHGDIYVNCFLAAVEPAPVLAAILVLOHLPRIYSALFLGSLVLFLOVPSLFLYSLALVWVGFITSASVWYVYTAIXPTVVRNMGVSSSTRSLGSLSPFYVLGAYDRFLPYILMGLTITLITLTFPFESGALPLPTIDMLRVKIRIKOMQDOSQIRYKDGESPTVAKSTAF"  
BASE COUNT 681 a 786 c 747 g 793 t  
ORIGIN  
Query Match 46.1%; Score 983.6; DB 10; Length 3007;  
Best Local Similarity 75.2%; Pred. No. 1.3e-178;  
Matches 1242; Conservative 0; Mismatches 404; Indels 6; Gaps 1;  
QY 126 ttcgagcgcagctgtggaagcattgcggagctacgacgaggtatgccttctctgagcag 185  
D6 66 TGTGTGGAAGCTGAGAGACGGCATGCGGAGCTACGAGAGGTGACGCGCTTCTGCGGAG 125  
QY 186 tgggggccccttcagcagcctatcttctctgtctcagcagcagcatatcccataatgac 245  
D6 126 TGGGGGCCCTTCCAGGCGCTCATTTTCTTCCGTCAGGCGCACGATCATCCCCATGCG 185  
QY 246 ttcattgtatgtcagctgtctctgtgcggagcccgagcagcagcagctgtcagatgacg 305  
D6 186 TTCAATGGTATGTCTATGCTGTCTGTGCGGGGAGACCCCGAGACACCGTGGCTTGTGCT 245  
QY 306 gacgcgcgaacctgagcagcctgtgcgaacaacagctgtccgcgtgcggtgcgcgagc 365

Db 246 CACACCGTGAACCTTTAGAGCAGCCGTGGGCAACCACTATCCCTTGGAGACGAGAC 305  
QY 366 ggcgcgaggtgcccccaagctgcagccgtactacggctgcgaaccatccgaactctcg 425  
Db 306 GGACGACAGGTGCTCACAAGAGTGCCTGCGCTACCGACTGCGCACACTGCGCAACTTCT 365  
QY 426 ggcctcgaggtggaacccggggcgagctggaacctggggcagctggaagcaagaagctgc 485  
Db 366 GCGCTGGGGCTGGAGCCCGGAGGAGCGTGAACCTGGAGACGTGGAGCAAGAGAACTGC 425  
QY 486 ctgagctggtggaagctcaagcaagcgtctaacctgtccacagctgcgtaccgaatggaat 545  
Db 426 CTGGATGGCTGGGAGTACACAGAGACGCTTCTGTCACACATCGTACAGAGTGGGAC 485  
QY 546 ctgctgtgtgaagacaactggaaggtgcacctacacacctccctgtctctcgtagagctg 605  
Db 486 CTGGTGTGTGAAGAGTGAAGTGGMAAGCCCACTACACACTCTCTGTTTGTGGGTTG 545  
QY 606 ctccctcgctcctctgtgtcggggagctgtcagacaggttggcagaagaagctctc 665  
Db 546 CTGATGGGCTCTTCAATTCGGGACAGCTCTCAGACAGGTTGGTGGCAAGATGTGCTG 605  
QY 666 ttgcgaacacatgctgtactacagactgtctcaagctctccctgcagatcttctccatcaagctg 725  
Db 606 TTTTTCACACATGGGATGACAGACTGGCTTCAAGCTTCTGCAAGCTTCTGTGAACCTTC 665  
QY 726 gaagatgtcactgtgtattgtatcagctggtggcatgggacagatcccaactatgtgta 785  
Db 666 GAGATGTTTACAGTCTCTTTGTGCTTGTGGCARGGTGACAGTCTCCACATAGTGGA 725  
QY 786 gctctcatcactagagacaagaatctctggaagctgactgctatatactctacatca 845  
Db 726 GCATTGTTCCTGGGAACGAATTTCTTCCAAATGCAATTCGAATTAATTTGGCCACCTTA 785  
QY 846 ggaagtgtcacaatttttgcagttggtctatagctgcgtccacactgttcttactatc 905  
Db 786 GGAGTTCATCATTTTATGCAATTTGGCTTCAATGGTGGCTGCTGTTGATCATCTTC 845  
QY 906 agaagctggcgagatgctgtctgtcgcgctgagcgtgcccggagagctgtgtgtcccgctg 965  
Db 846 AGAGACTGGAGATGCTCTGCTGGCTCACTGTACATGAGGGGGTGTGGGGCTCTTC 905  
QY 966 tggttgttcatcctcgaatcccccgaatgcgtgatatcccccgaagaattagaagagct 1025  
Db 906 TGGTGTTCATCCCTCGAGTCCGCCGATGCGTCACTCTCAAGGCCGAGTTAAAGAGCA 965  
QY 1026 gaagatatcatccaaaagctgtcaaaaatgaacaacacagctgtacagcagtgatatt 1085  
Db 966 GAGGTGATCATCCGCAAGAGCTGCCAAATTCATGGGATTTGGGCTTCCACTATCTTC 1025  
QY 1086 gatctgtggag-----gaactaatccctgaagcagcagaagaagcttattatctgagc 1139  
Db 1026 GATCCCACTAGGTACAAAGACTTAAATTCCAAGAGCCTCGAGTCGCACCACTAATTAAT 1085  
QY 1140 ctgttcgaagactcgaatatcgtccataatgaacatlatgctctgtcgtcctatgagatcgt 1199  
Db 1086 CTGATCGAACAAGAAATATCAGATCATCAACATCAATGCTCATATCTGTGCTGACC 1145  
QY 1200 acctcagtggttactgtctgtctgtctgtcgtcctcctaaattcaatgagagatgcctac 1259  
Db 1146 ATATCAGTGGGCTACTTGGACTTCTCTGTACACTCTAATCTGATGGGAGCATCTAT 1205  
QY 1260 ctgaagctttcctcctcgtcgtctgattgaattccagcttaccatcaagcctgctgcta 1319  
Db 1206 GTGAATCTCTTCTACTGTGGAGCTGTGAATCCCGCTATAGTGGCTGTGCTGTG 1265  
QY 1320 ttgcgaagcgtgcgccagcgtlatatcatatcagctcgaagctgtctcgtggagaggtgtg 1379  
Db 1266 CTACAGCAACCTGCCCCGGGATTCATCTCTGCTGCCCTTCCGTTGGTGGAGAGTGC 1325  
QY 1380 ctctctctcatcgaactggtgtaacctgtggaattattatccatctgactgtgcatg 1439  
|||||

Db 1326 CTTCCTTCATACAGCTGCTGCTTCAGATTTGTTTACTTGTCCACTGCCCTGGTGATG 1385  
QY 1440 ctgggaaatttggatcacctctgcttctccatgctgtatgtctctcctgagctc 1499  
Db 1386 GTGGGCAAGTTTGAATACACTCTGCTTATTCATGCTATGTACACGGCCGAGCTG 1445  
QY 1500 taccacaacctgtgtcgaagaacatgagcggtgtggtgtacatccacggctccagagtgagc 1559  
Db 1446 TACCCCACTGTGTGCAAAAACATAGGTGTGGGTGAGCTCAGACATGCCCTTGGC 1505  
QY 1560 agcatatgccccctactctgttactcctcgtggtcttacaacaagaatgtgcctcatc 1619  
Db 1506 AGCATCTGTCTCCACTTGTCTACTGTGTGCTATGATGCTTCCTGCTTATATC 1565  
QY 1620 gtaatggtagtctgactgtctcgtatgtgaatcttcaaccttlttctcctgaagctg 1679  
Db 1566 CTATGGGGAAGTGTGACATCTGACAGCATCTCACTGTTGTTCTCCAGAGACTTT 1625  
QY 1680 ggaatgactcttccaagaacttagagcagatgcagaagaatgagtgatcagatcgg 1739  
Db 1626 GTTCCCTCTCCCGACACCATTTGACCAATGCTCAAGGCTCAAGAAATAAACAATG 1685  
QY 1740 aaaaaacaagagactcaatgtgagacagaaga 1771  
Db 1686 CAATCCAAAGCCAAACAGAGCAAAAAGA 1717  
  
RESULT 13  
LOCUS AF110416 3037 bp mRNA linear ROD 10-SEP-1999  
DEFINITION Rattus norvegicus organic cation/carnitine transporter (OCTN2)  
ACCESSION AF110416  
VERSION AF110416.1 GI:5852403  
KEYWORDS Norway rat.  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 3037)  
AUTHORS Wu,X., Huang,W., Prasad,P.D., Seth,P., Rajan,D.P., Leibach,F.H.,  
Chen,J., Conway,S.J. and Ganapathy,V.  
TITLE Functional characteristics and tissue distribution pattern of  
organic cation transporter 2 (OCTN2), an organic Cation/Carnitine  
transporter  
JOURNAL J. Pharmacol. Exp. Ther. 290 (3), 1482-1492 (1999)  
MEDLINE 99384224  
PUBMED 10454528  
REFERENCE 2 (bases 1 to 3037)  
AUTHORS Wu,X., Huang,W., Prasad,P.D., Rajan,D.P., Leibach,F.H.,  
Conway,S.J. and Ganapathy,V.  
TITLE Direct Submission  
JOURNAL Submitted (01-DEC-1998) Department of Biochemistry and Molecular  
Biology, Medical College of Georgia, 1120 15th Street, Augusta, GA  
30912-2100, USA  
  
FEATURES  
source  
1..3037  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
1..3037  
/gene="OCTN2"  
101..1774  
/gene="OCTN2"  
/codon\_start=1  
/product="organic cation/carnitine transporter"  
/protein\_id="AAD54059.1"  
/db\_xref="GI:5852404"  
/translation="MRDYDEVATYRILEGEMPRORLFFLLSASIIIPNGFNGMSIVFLAG  
TEHRCIVPHRYNLSASARNHNSPILETMDGROVRSCKRYRLATIANRNSALCEPGRD  
VDEQLEDENCLDGEYNKDYFLSTIVTBMIDLVCKDDKAPLITSLFVGVGLMSFIS  
GQLSDRFGRKNVLFITMGQMGTGFSFLQLEFVNFETVLFVLVVGQILSNVVAFLVIG  
TEILSKSIRIIFATLGVCIFFYAFGFMVPLPFAFYTRDMRLALITVPGVLCALMWF









RESULT 15  
AB018436  
LOCUS AB018436 2297 bp mRNA linear ROD 21-DEC-2000  
DEFINITION Mus musculus mRNA for OCTN3, complete cds.  
ACCESSION AB018436  
VERSION AB018436.1 GI:4996130  
KEYWORDS OCTN3.  
SOURCE Mus musculus adult testis cDNA to mRNA.  
ORANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;  
1 (sites)  
REFERENCE  
AUTHORS Tamai, I., Ohashi, R., Nezu, J.I., Sai, Y., Kobayashi, D., Oku, A.,  
Shimane, M. and Tsuji, A.  
TITLE Molecular and Functional Characterization of Organic  
Cation/Carboline Transporter Family in Mice  
JOURNAL J. Biol. Chem. 275 (51), 40064-40072 (2000)  
PUBMED 11010964  
REFERENCE 2 (bases 1 to 2297)  
AUTHORS Nezu, J.  
TITLE Direct Submission  
SUBMITTED (07-OCT-1998) Jun-ichi Nezu, Chugai Research Institute  
for Molecular Medicine, Inc., Gene Search Program: 153-2 Nagai,  
Nihari, Ibaraki 300-4101, Japan (E-mail:nezu@cimmed.com,  
Tel:81-298-30-6211, Fax:81-298-30-6270)  
FEATURES  
SOURCE location/Qualifiers  
1..2297  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/tissue\_type="testis"  
/dev\_stage="adult"  
171..1865  
/gene="OCTN3"  
171..1865  
/gene="OCTN3"  
/note="Carnitine transporter in testis"  
/codon\_start=1  
/product="OCTN3"  
/protein\_id="BA078343.1"  
/db\_xref="GI:4996131"  
/translation="MLDYDEVTAFLGEMGTFORLIFELLSASIIIPNGFTGLSAVFLTA  
IPEHRRIPTVNLSSAMRNHSIPMETKDEPPOKRRRLATIANFSELGEPGD  
VDLEOLBOENCLDMEYDKDIFLSTIVTENDLVCKDKMKAPLTSPFYVGLGFS  
GOLSDRGRKNILFLFMAHMGFSFIOFVSNEFMPLLYTVMGHSIYVAAPVIG  
TEMKSVRIITFLGICIFPARGFEMVLPFAFIRMRRLALATIPVLCALMMF  
IPESRNLISOGRIKEAEVITIRAKINGIVASTIFDPSETKLDDSKKQSHHI  
YDIVRTNPRIRLITMSILMLTISVGFGLSDTPNNGNIYVNCFLAIVAPVYL  
AMLIQHSRYRSMAGSLFLGGSVLLVQVPSDPLNGLIYVNGFGLTSAVAVL  
VYTAELIPTVARNMGVSVSTASRLGISLSPFYGLCAVDRPPIYIMGSLITLTAI  
TLPPESSGSLPETIDEMQVKRLKROSLSKRGSKESKGNVSRSRSEPKGF"  
BASE COUNT 523 a 618 c 571 g 584 t 1 others  
ORIGIN

Query Match 42.4%; Score 905.6; DB 10; Length 2297;  
Best Local Similarity 72.8%; Pred. No. 1,1e-163;  
Matches 1188; Conservative 0; Mismatches 429; Indels 15; Gaps 1;

QY 134 ggcacatgggaagcagcgagactacgacgaggtgacgccttcctggcgagtgaggcc 193  
DB 158 GGCTGCCGACCTGATGCTTGAAGACGAGGTGACCCGCTTCGCGGAGTGGGAGC 217

QY 194 cttcagcgccctcatctcttcctgctcagcgccgagcctatcccaatgctcaagg 253  
DB 218 CTTCCAGCGCCATCTCTCTCTCTCAAGCCGACATCATCCCAATGAGCTTACTGG 277

QY 254 tatgtcagtcgtgtctctcgtcgcggaagcccgagacgcgtctcagtgccgagccgc 313  
DB 278 GTTGTAGCGCGTCTTCTCTAGCGCGATGCCGAGACACCGTTCCGGATACAGAACCGT 337

QY 314 gaacctgagcagcgcctcgtgcgaacaacagtgctccgcgtcggtcgtcgagccgcga 373  
DB 338 GAACCTGACAGCGCGTGGCGCAACCAAGTATCCCGATGGAGACGAGGAGCCAGCA 397

QY 374 ggtgccccagctgagcagcgcgtacagcgtcgcacccatccgcgaacttcgtcgctgcg 433  
DB 398 GGTGCTCGAAGTATGCGCGCGCTACCGAGTGGCCACATCGCCAACTTCTGAGCTGGG 457

QY 434 gctcgagccgagcgagcgtgagcctcgtgagcagcgtgagcagcagcgtcgtcgtatgg 493  
DB 458 GCTGGAGCGCGGAGCGGAGCGTGTGACCTGGAGACGCTGGAGCAAGAACTGCTGGATGG 517

QY 494 ctggaggttcaagcagcagcgtctacctgtccacgcgtcgtgaccgagtggaatcgtgtg 553  
DB 518 CTGGGAGTACGACAGACATCTTCTGTCCACATCGTACAGAGTGGAGCTGGTGTGG 577

QY 554 tggagcaactggaaggtgccccctacacacccctccgtctcgttagggcgtcctcgtg 613  
DB 578 TAGGATGACTGGAAGCCCACTACACACTCTCTTTACGTGGGTGTCTTTAGG 637

QY 614 ctccctcgtgtccggcgacgctgcagacaggttgcgaggaagacgtctcttcgcaac 673  
DB 638 CTCCTTCATTTGCGGACAGCTCTCACACAGGTTTGGTCCAGAAATCTTGTTTTGGAC 697

QY 674 catgctgtacagcagcgtcagctcagctcctcgcagatttctccatcagctggagatgt 733  
DB 698 CANGGCATCACACCGGATGAGCTTACGTTACAAAGCTCTCTGTGAACTTGAGATGTT 757

QY 734 cactgtgtattgtatcgttggtgacatgggcaagctcccaatcagtgtagtgcctcat 793  
DB 758 TACTCTGCTCTATACCTCTGTTGAAATGGGACATATATCACTACGTGGACGATTTGT 817

QY 794 actaagagcaagaactcttcgtcagcagcagctgtctatatactctacatgagatgt 853  
DB 818 CTTGGGACAGAAATGCTTCCAAAGTCAGTATGAATATATTCGCCACTTGGAGATTTGG 877

QY 854 cacattttgtcagctgtgcctatagctcgtgcacacgtgttgccttactcatcagagactg 913  
DB 878 CATATTTTTCGCTTGGCTTGCATGTGCTGCGCTCTGTTTGCATCTATCATGAGAGTGT 937

QY 914 gcgagatgctcgtcgtcggtgacggtgcgcgagagcgtgtgtccgcgtgtgtgtgt 973  
DB 938 GAGGAGGCTGCTGTGGCAATTACTTTACAGGCGTGTGTGGGGCTCTGGGTGTT 997

QY 974 catccctgaatctcccgatgctgatatcccgagaagaatattagaagcgtgaagatat 1033  
DB 998 CATCCCTGAGTCCCGCAGATGAGCTCATCTCTCAAGGCCGCAATTAAGAGAGGAGGTGAT 1057

QY 1034 catccaaaagcgtgcaaaaatgaacaacacagcgtctgacagcagtgatatattgattcgt 1093  
DB 1058 CATCCCGCAAGCGCCCAAAATCAATGGGATTTGTGGACCTTCCACATATCTTGATCCAAAG 1117

QY 1094 gggg-----gagctaaatcccccgaagcagcagaagaagcttcaatcttgga 1138  
DB 1118 CGAGACCAATTAATTTACAGACATAGTTTCAAGAGAGCCCACTGCGACCACTTATATGA 1177

QY 1139 cctgttcagcagctcgaataattgcataatgacatattgcttctgtcgtatgagatgt 1198  
DB 1178 TCTGTGTCGAAGCAACCAATATACAGATCCCTACACATCATGCTATATCTCTGTGCTGAC 1237

QY 1199 gaacctgagtggttacttctgtcctcgtcgtgagctcctcaattacatgagagatgctcta 1258  
DB 1238 CATATCAGTGGGCTATTTTGAAGTATCTTTGACACTCTTCACTTAACTTAAGGGAACATCA 1297

QY 1259 cctgaactgttctcctcctcgtcgttgaattgaattcagcttcaatcagcagcgtgctgt 1318  
DB 1298 TGTGAACCTGCTTCTACTGTGGCGGCTGTGAAGTCCAGGCTTATGTCTGGCTGGCTGTT 1357

```
QY 1319 attgcgaacgctgcggcggtatatacatagctgaagtaactgtctcgtggaggaagtgct 1378
    |||||  |||  |||||  |||  |||  |||||  |||  |||  |||||  |||  |||
Db 1358 GTTGCAGCATGTGTCCGGCGGTATTCTATGGCTGGTTCCTCTCCGTGGGCGAGTGT 1417
    |||||  |||||  |||||  |||  |||  |||||  |||  |||  |||||  |||  |||
QY 1379 gcttcctcaatcaactcgtacactcgtgattattactcttatccaatcgtcgtcat 1438
    |||||  |||||  |||||  |||  |||  |||||  |||  |||  |||||  |||  |||
Db 1418 CCTTCTTATGACAACTGTGCTCAAGACCTACATTACTGTCTACTACCTGTGAT 1477
    |||||  |||||  |||||  |||  |||  |||||  |||  |||  |||||  |||  |||
QY 1439 gctgggaaatttggaatacctcgtcttcctcactcgtcgtatcctcgaactcgaact 1498
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1478 GGTGGGGAAGTTGGATACCTTGCTGCTACTCCATGGCTATGTGTACACAGCTGAGCT 1537
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1499 ctaccacaacctcgtcagaacatgacgtggggg lcaacatcacgacctcgaagtgg 1558
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1538 GTACCCCACTGTGTGAGAAACATGGGTGTGGGGGTCAAGCTCCACAGCATCCCGCCTTGG 1597
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1559 cagcatcatcgcacctactttgtttactcgtcgtcttaacaagaatcgtgcctacat 1618
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1598 CAGCATCCTGTCTCCCTACTTGTGTATGTAGTGCCTATGACCGCGCCTACCTTATAT 1657
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1619 cgtcatgggtaagtcgaactcgtcgtatggaaatcttcaaccttttccctgaaagtt 1678
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1658 CCTCATGGGAAGTCTAACCATCTGACAGCTATCATCATTTATCTTCCAGAGAGTTTC 1717
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1679 ggggaatgactctccaagaaaccttagagcagaatgcagaagtgaatcgtcgtcgt 1738
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1718 TGGGGTTTCTCTCCAGAAACCATTTGACGAGATGCAGAAAAAGTCAAAAAAAGTAAACAAAG 1777
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1739 gaaaaaaacaag 1750
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1778 GCAATCCCTAAG 1789
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: July 16, 2002, 23:46:19  
Job time: 15010 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2002, 20:37:14 ; Search time 596.09 Seconds  
(Without alignments)  
6149.428 Million cell updates/sec

Title: US-09-521-195b-2

Perfect score: 2135

Sequence: 1 cccgcgcttcgcgcgcccaat.....aatactatccaataaataat 2135

Scoring table: IDENTITY\_NUC

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq\_032802:\*

1: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA2001.DAT:\*

23: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SIDB5/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2135	100.0	2135	20	AAAX26879
2	1310.6	61.4	2083	20	AAAX26898
3	1066.4	49.9	1831	20	AAAX26880
4	1066.4	49.9	1831	21	AAA09889
5	1066.4	49.9	3261	23	AAAS67216
6	994	46.6	1888	20	AAAX26902
7	994	46.6	1888	21	AAA09890
8	905.6	42.4	2297	21	AAA86053
9	505	23.7	700	22	AAH92757

C	10	363.8	17.0	700	22	AAH92241	Human inflammatory
C	11	363.8	17.0	25871	21	AAA09888	Human genomic OCTN
C	12	314	14.7	700	22	AAH92293	Human inflammatory
C	13	297	13.9	700	22	AAH92294	Human inflammatory
C	14	234	11.0	700	22	AAH92756	Human inflammatory
C	15	217.6	10.2	700	22	AAH92677	Human inflammatory
C	16	193.6	9.1	2402	22	AAK52384	Human polynucleotl
C	17	189.6	8.9	2533	23	ABL10925	Drosophila melanog
C	18	189.6	8.9	4533	23	ABL10924	Drosophila melanog
C	19	187.8	8.8	2070	21	AACT7247	Human ORFX ORF2802
C	20	187.8	8.8	2110	22	AAK51500	Human polynucleotl
C	21	184.4	8.6	1413	23	AAV73319	DNA encoding novel
C	22	184.4	8.6	1662	19	AAV49559	Human liver cell c
C	23	184.4	8.6	1888	19	AAV49558	Human liver cell c
C	24	184	8.6	700	22	AAH92675	Human inflammatory
C	25	176	8.2	700	22	AAH92702	Human inflammatory
C	26	175	8.2	1882	17	AAH08702	Rat OCT-1 gene. R
C	27	165.2	7.7	1901	21	AAZ44679	Rat liver anion tr
C	28	158.4	7.4	700	22	AAH92719	Human inflammatory
C	29	150.8	7.1	1865	22	AAH90101	Human bone marrow
C	30	150	7.0	700	22	AAH92210	Human inflammatory
C	31	147	6.9	2054	22	AAH90049	Human bone marrow
C	32	144.6	6.8	2087	23	ABL23075	Drosophila melanog
C	33	144.6	6.8	4087	23	ABL23074	Drosophila melanog
C	34	143.6	6.7	1950	22	AAAC85821	HOAT2B DNA. Homo
C	35	143.2	6.7	2501	22	AAAC85820	HOAT2A DNA. Homo
C	36	140.2	6.6	700	22	AAH93273	Human inflammatory
C	37	136.4	6.4	1697	22	AAAC83979	Murine organic ani
C	38	135.8	6.4	2179	21	AAAL1146	Human cerebral org
C	39	134.2	6.3	2121	22	AAAC85822	HOAT3 DNA. Homo s
C	40	133	6.2	700	22	AAH93282	Human inflammatory
C	41	133	6.2	3009	23	AAH93888	DNA encoding novel
C	42	132	6.2	700	22	AAH92676	Human inflammatory
C	43	129	6.0	3229	23	AAH75476	DNA encoding novel
C	44	128	6.0	3229	23	AAH75478	DNA encoding novel
C	45	127.8	6.0	2135	18	AAV12393	Human osteoclast t

#### ALIGNMENTS

RESULT	1
ID	AAAX26879 standard; DNA: 2135 BP.
AC	AAAX26879;
DT	23-JUN-1999 (first entry)
DE	DNA encoding a protein with cation transporting activity.
KW	Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
KX	heart disease; cancer; anti-tumour drug; anticancer drug; ss.
OS	Homo sapiens.
FT	Key CDS
FT	Location/Qualifiers
FT	147..1802
FT	/*tag= a
PN	W09913072-A1.
PD	18-MAR-1999.
PF	07-SEP-1998; 98WO-JP04009.
PR	20-MAY-1998; 98UP-0156660.
PR	08-SEP-1997; 97JP-0260972.
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI	Nezu J, Oku A;

DR WPI: 1999-215062/18.  
DR P-PSDB: AAY01649.  
PT Genes homologous with organic cation transporters OCT1 and OCT2,  
PT useful in design of new drugs for treatment of diseases due to  
PT abnormality of the transporter functions  
XX  
PS Claim 2; Page 45-51; 97pp; Japanese.  
XX  
CC The present sequence encodes a protein with cation transporting  
CC activity. The genes are significantly homologous with organic cation  
CC transporters OCT1 and OCT2. The genes may used in drug development,  
CC particularly in the treatment of diseases due to abnormality of the  
CC organic cation transporter functions e.g. fatty liver, heart diseases  
CC and cancers, by controlling such as by inhibition or activation.  
CC Administration of anti-tumour and anticancer drugs in combination with  
CC a transporter protein inhibiting agent allows the agents to penetrate  
CC into the diseased cells to enhance the drug action.  
XX  
S0 Sequence 2135 BP; 499 A; 547 C; 530 G; 559 T; 0 other;

Query Match 100.0%; Score 2135; DB 20; Length 2135;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ccccgcttcgcccccaattcttaacagcctctgtcccccgggaagcttcaacatc 60  
Db 1 ccccgcttcgcccccaattcttaacagcctctgtcccccgggaagcttcaacatc 60  
QY 61 ctggggagggccccaagtacaagaacactgtccctggaagaagcttcaacccgtagt 120  
Db 61 ctggggagggccccaagtacaagaacactgtccctggaagaagcttcaacccgtagt 120  
QY 121 caaatttcggagcgagcagttgggaagcaltgcggaactagagaaagtatcgcttcctg 180  
Db 121 caaatttcggagcgagcagttgggaagcaltgcggaactagagaaagtatcgcttcctg 180  
QY 181 gcgagtgggggccctcccaagcctcatcttctctgtctgaagccagatcatcccca 240  
Db 181 gcgagtgggggccctcccaagcctcatcttctctgtctgaagccagatcatcccca 240  
QY 241 atggtcctcaatgtatgtatcagtcgttctctgcgggagcccggaacccgtctcgag 300  
Db 241 atggtcctcaatgtatgtatcagtcgttctctgcgggagcccggaacccgtctcgag 300  
QY 301 tgcggagcgcgcggaacctgagcagcgctgcgcaacaacagtgctcccgctgcgctgc 360  
Db 301 tgcggagcgcgcggaacctgagcagcgctgcgcaacaacagtgctcccgctgcgctgc 360  
QY 361 gggagcgcgcgcggaagtgcccccaagctgacagcgcgtacccggtctcgccacatcgcaact 420  
Db 361 gggagcgcgcgcggaagtgcccccaagctgacagcgcgtacccggtctcgccacatcgcaact 420  
QY 421 tctcgagcctcggtctggaagcgggagcagctggaacctggagcgggagctggagcaggaga 480  
Db 421 tctcgagcctcggtctggaagcgggagcagctggaacctggagcgggagctggagcaggaga 480  
QY 481 gctgcctggaatggtggaattcaagccaggaagctactactgtccacacgtctgtagccagagt 540  
Db 481 gctgcctggaatggtggaattcaagccaggaagctactactgtccacacgtctgtagccagagt 540  
QY 541 ggaatcgtgtgtgtggaagaacttggaagtgccctctcaacacactccctgtctctctag 600  
Db 541 ggaatcgtgtgtgtggaagaacttggaagtgccctctcaacacactccctgtctctctag 600  
QY 601 ggcgtgctcctcgctcctctgttccgggagcagctgtccagacaggtttggcaggagaagag 660  
Db 601 ggcgtgctcctcgctcctctgttccgggagcagctgtccagacaggtttggcaggagaagag 660  
QY 661 ttctcttcgaaacatagctgtacagactggtctcagcttccctgcaagattttctccatca 720  
Db 661 ttctcttcgaaacatagctgtacagactggtctcagcttccctgcaagattttctccatca 720

QY 721 gctgggagagttccactggttattgtcatgctgtgggcatgggccaagatccccaactatg 780  
Db 721 gctgggagagttccactggttattgtcatgctgtgggcatgggccaagatccccaactatg 780  
QY 781 tggtagcctcatactaggaacagaattcttggcaagtcagttgatattctctca 840  
Db 781 tggtagcctcatactaggaacagaattcttggcaagtcagttgatattctctca 840  
QY 841 catlaagaagtgtcaaatlttttgcagttggtatatagtcgtccacagttgttact 900  
Db 841 catlaagaagtgtcaaatlttttgcagttggtatatagtcgtccacagttgttact 900  
QY 901 tcatcagaagcctggcgatgtgctgtcgctgagcagtgccgggaagtgctgtgtcc 960  
Db 901 tcatcagaagcctggcgatgtgctgtcgctgagcagtgccgggaagtgctgtgtcc 960  
QY 961 cgcgtgtgtgtcatctccgaatcccccgaatggtgtgatataccagagaagatttagag 1020  
Db 961 cgcgtgtgtgtcatctccgaatcccccgaatggtgtgatataccagagaagatttagag 1020  
QY 1021 aggcctgaagatatacatccaaaagctgcaaaatatgacaacacagctgtaccagcagtg 1080  
Db 1021 aggcctgaagatatacatccaaaagctgcaaaatatgacaacacagctgtaccagcagtg 1080  
QY 1081 tattgatctgtggagggagctaaatcccttgaaagcagaagaagcttcatcttgagacc 1140  
Db 1081 tattgatctgtggagggagctaaatcccttgaaagcagaagaagcttcatcttgagacc 1140  
QY 1141 tgttcagagctcggaaatatgtccataatgacatlatgtcttctgtctatgtatgtcga 1200  
Db 1141 tgttcagagctcggaaatatgtccataatgacatlatgtcttctgtctatgtatgtcga 1200  
QY 1201 cctcaatgggttaacttgtctgtctgtctgtatgtctccaaattacatggagatgctacc 1260  
Db 1201 cctcaatgggttaacttgtctgtctgtctgtatgtctccaaattacatggagatgctacc 1260  
QY 1261 tgaacgttctctctcctgctcgatgtgaattccagcttaacatlaacagcctgtgcgtat 1320  
Db 1261 tgaacgttctctctcctgctcgatgtgaattccagcttaacatlaacagcctgtgcgtat 1320  
QY 1321 tgcgaagcgtcccccagcggttatatcatagctgcagactgtctctgggaggaagtgctgc 1380  
Db 1321 tgcgaagcgtcccccagcggttatatcatagctgcagactgtctctgggaggaagtgctgc 1380  
QY 1381 tctcttcaatcaactgtggaacctgtggaattatattacttttccatttgcgtgtcaagc 1440  
Db 1381 tctcttcaatcaactgtggaacctgtggaattatattacttttccatttgcgtgtcaagc 1440  
QY 1441 tgggaanaattgggaatcaacctctgtcttccatgtcttatgtcttcaactgtcagctct 1500  
Db 1441 tgggaanaattgggaatcaacctctgtcttccatgtcttatgtcttcaactgtcagctct 1500  
QY 1501 acccaacctgtgtcaggaacatggtcggtgtggggttcaatccacgggctccagagtgggca 1560  
Db 1501 acccaacctgtgtcaggaacatggtcggtgtggggttcaatccacgggctccagagtgggca 1560  
QY 1561 gcatcatggccccctacttgtttacccctgggtgtcttaacaaagaagtgtgccccactcg 1620  
Db 1561 gcatcatggccccctacttgtttacccctgggtgtcttaacaaagaagtgtgccccactcg 1620  
QY 1621 tcatgtggtatctgagctgtcctgtatgtgaatcttcaacccttttccctggaaggtttgg 1680  
Db 1621 tcatgtggtatctgagctgtcctgtatgtgaatcttcaacccttttccctggaaggtttgg 1680  
QY 1681 gaatgactcttccgaagaacctttagagcagatgacagaaagtgaatgttcagatctggga 1740  
Db 1681 gaatgactcttccgaagaacctttagagcagatgacagaaagtgaatgttcagatctggga 1740  
QY 1741 aaaaacaagaagactcaatgtagagcagaagaataatcccaaggttctaataactgcatctc 1800  
Db 1741 aaaaacaagaagactcaatgtagagcagaagaataatcccaaggttctaataactgcatctc 1800

QY	1801	gaaaaaatcttaccocattggtggaagfgaaaaaagaataataagacctgtgaggaa	1860
DB	1801	gaaaaaatcttaccocattggtggaagfgaaaaaagaataataagacctgtgaggaa	1860
QY	1861	attcgttcttccacctaagaatgagcgtactgtgaacgatgtgacaccaaatagaacctgct	1920
DB	1861	attcgttcttccacctaagaatgagcgtactgtgaacgatgtgacaccaaatagaacctgct	1920
QY	1921	atcaagaaatgctgtcatcaatgaactctgagatattcttcacgataaagtctctgct	1980
DB	1921	atcaagaaatgctgtcatcaatgaactctgagatattcttcacgataaagtctctgct	1980
QY	1981	ttacaacaacaacctttctcgaagagctcctctactcatatattcaatgaaatgagttgct	2040
DB	1981	ttacaacaacaacctttctcgaagagctcctctactcatatattcaatgaaatgagttgct	2040
QY	2041	aagatgctctggaanaactgttaagcaagagcgtgtaaaataatataagtttaaacctc	2100
DB	2041	aagatgctctggaanaactgttaagcaagagcgtgtaaaataatataagtttaaacctc	2100
QY	2101	atttccaatcatatacatatactatccaataaanaat	2135
DB	2101	atttccaatcatatacatatactatccaataaanaat	2135

RESULT	2	
AXX	AXX26898	
ID	AXX26898	standard; DNA; 2083 BP.
AC	AXX26898;	
XX		
XX	23-JUN-1999	(first entry)
DT		
DE	DNA encoding a protein with cation transporting activity.	
XX		
XX	Organic cation transporter; OCT1; OCT2; drug development; fatty liver;	
KW	heart disease; cancer; anti-tumour drug; anticancer drug; ss.	
XX		
OS	Mus musculus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	122..1783
FT		/*tag= a
XX		
PN	W09913072-A1.	
XX		
PD	18-MAR-1999.	
XX		
PE	07-SEP-1998;	98WO-JP04009.
XX		
PR	20-MAY-1998;	98JP-0156660.
PR	08-SEP-1997;	97JP-0260972.
XX		
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.	
XX		
PI	Nezu J, Oku A;	
XX		
DR	WPI: 1999-215062/18.	
DR	P-PSDB; AAY01651.	
XX		
PT	Genes homologous with organic cation transporters OCT1 and OCT2,	
PT	useful in design of new drugs for treatment of diseases due to	
XX	abnormality of the transporter functions	
XX		
PS	Claim 2; Page 68-74; 97pp; Japanese.	
XX		
CC	The present sequence encodes a protein with cation transporting	
CC	activity. The genes are significantly homologous with organic cation	
CC	transporters OCT1 and OCT2. The genes may used in drug development,	
CC	particularly in the treatment of diseases due to abnormality of the	
CC	organic cation transporter functions e.g. fatty liver, heart diseases	
CC	and cancers, by controlling such as by inhibition or activation.	
CC	Administration of anti-tumour and anticancer drugs in combination with	

CC a transporter protein inhibiting agent allows the agents to penetrate  
CC into the diseased cells to enhance the drug action.  
XX  
SQ Sequence 2083 BP; 477 A; 549 C; 548 G; 509 T; 0 other;

Query Match	61.4%;	Score 1310.6;	DB 20;	Length 2083;
Best Local Similarity	85.7%;	Pred. No. 0;		
Matches 1470; Conservative	0;	Mismatches 239;	Indels 6;	Gaps 1;

[illegible]



```

Db 582 gctgttggtcccttcatttcagggcagctgtcacagagtttggccggagaagtgtct 641
QY 665 ctccgaacatggtcgtacagactggtcgtccctctcgagatttctccatcagtct 724
    |||| ||||| | |||| | ||||| ||||| ||||| ||||| ||||| |||||
Db 642 gtctgtgacacagggcagcagacagcgtccctccgcagctctctcgaaagattc 701
QY 725 ggaagatgtcactgtgtatttgcattgtgagcattggcagacatccaaatagtgt 784
    |||| ||||| | |||| | ||||| ||||| ||||| ||||| ||||| |||||
Db 702 tgagatgttgcgtgtgtgtctgtctgtgagcattggccagatcccaactatgtgc 761
QY 785 agcctcatcaggagaagaagaattctgtgcagatcatttgtattatctctcaatt 844
    |||| | | | | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 762 agcattgtccgggagagaagaattctgtgcagatcatttgtattatctctcaatt 821
QY 845 aggaatgtgacatttctgtgattgtgtatgtgtgtgcacgtgttcttctctcat 904
    ||||| ||||| |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 822 aggaatgtgacatttctgtgattgtgtatgtgtgtgcacgtgttcttctctcat 881
QY 905 cagaagactggcggatgctgtgtgtgtgcgtgcagcgtgcgggagtgctgtgtccgct 964
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 882 ccgaagactggcggatgctgtgtgtgtgcgtgcagcgtgcgggagtgctgtgtccgct 941
QY 965 gtgtgtgtcattcccttcgattcccccggatggcgtatcccaagagaatttaagagc 1024
    ||||| ||||| |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 942 ctgtgtgtcattcccccggatggcgtatcccaagagaatttaagagc 1001
QY 1025 tgaagatataccaagaagctgcaaaatgacaacaacagctgtaccagagtgatatt 1084
    || | ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 1002 aagagtgatataccgcaagagctgcgaagacaaatgagatgtgtgtcttccactatctc 1061
QY 1085 tgattctgtgag-----gagctaaatcccttgaagcagagaagaacttcatcttga 1138
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1062 tgaccggatgagttacacaagaacttaagttccaagaagcagagcccaacaacttcgga 1121
QY 1139 cctgttcagagactcggaatatgtccataatgacataatgtctgtgtcatgagct 1198
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1122 tctgtcttcgaaccggaaatataccggaatggtacacataatgtccaaatgtgtgtgtgac 1181
QY 1199 gaactcagtggttaacttctgtctgtctcgtgagatgctcctaatttaagtgaatgacta 1258
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1182 catatcagtggtctatttctgtgtctgtgtatctcctaacttgaatggggagacatct 1241
QY 1259 cctgaactgttctctctgtctgtgtgtgaatccacagcttaacttaacagctgtgtct 1318
    ||||| ||||| |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 1242 tctgaactgtctctctccttcaggaatggttgaagtcacacatattgtgtgtgtgtgtct 1301
QY 1319 atgtcgaagctgcggcagcgttatatcatagctgcagtaactgtctggggagagagtgat 1378
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1302 gctgcaatatgtgcccggcgctattatcatagcgcacatgcccctctctctgtgtgtgtgt 1361
QY 1379 gcttctcttaactaactgtgacctgtgtgtattacttctatccatttgtgtgtcat 1438
    ||||| ||||| |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 1362 ccttctcttaactaactgtgacctgtgtgtattacttctatccatttgtgtgtgtcat 1421
QY 1439 gctggagaatttgggataccctgtcttctcactgtgtgtgtgtgtgtgtgtgtgtgtgt 1498
    | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 1422 ggtgtggcaagtgttggagtcacagctgtcttctcactgtgtgtgtgtgtgtgtgtgtgt 1481
QY 1499 ctaccacaacctgtgtcaggaacatgctgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1558
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1482 gtatccacaagtggtgagaacatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1541
QY 1559 cagatataatgcccactacttgttactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1618
    ||||| ||||| |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 1542 cagatataatgcccactacttgttactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1601
QY 1619 cgtcatagtgatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1678
    ||||| ||||| |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 1602 tctcatatggaagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1661
QY 1679 gggagatgactcttcacagaacactttagagcagatgcaagaatgtaattgtgtcagatctgt 1738
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1662 cgttaccacactccacagacacatgtgacagatgtaagtgtaagaatgtgaacaacag 1721

```

```

QY 1739 gaaaaaacaagactcaatgtgagacagaagaatc 1776
    |||| |||| | | | | | | | | | | |
Db 1722 aaaaactccaagtcacacaagaatgttaaaagtgtc 1759

RESULT 4
ID AAA09889 standard; cDNA; 1831 BP.
XX
AC AAA09889;
XX
DT 05-JUL-2000 (first entry)
XX
DE Human OCTN2 cDNA sequence.
XX
KW Organic cation transportation; human; carnitine transporter protein;
KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
KW Juvenile visceral steatosis; ss.
XX
OS Homo sapiens.
XX
PN WO200014210-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-JP04853.
XX
PR 07-SEP-1998; 98JP-0252683.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Nezu J, Oku A.
XX
DR WPI; 2000-256966/22.
DR P-PSDB; AAY83929.
XX
PT Systemic carnitine deficiency gene OCTN2 encoding part of organic
PT cation transporter, useful as diagnostic tool
PS Example 1; Page 41-46; 106pp; Japanese.
XX
CC This sequence represents the cDNA encoding the human carnitine
CC transporter protein OCTN2. The sequence can be used as a target for
CC diagnosis of systemic carnitine deficiency by detecting the presence
CC of mutations in the sequence, especially seen in the disease juvenile
CC visceral steatosis (jvs). The wild type OCTN2 gene can be used in the
CC gene therapy of the disease state.
XX
SQ Sequence 1831 BP; 348 A; 527 C; 513 G; 443 T; 0 other;

Query Match 49.9%; Score 1066.4; DB 21; Length 1831;
Best Local Similarity 78.2%; Pred. No. 3.7e-265;
Matches 1296; Conservative 0; Mismatches 356; Indels 6; Gaps 1;

QY 125 ttccgagagggcaggtgggagagatcgcggactacagagtgatgcctctcctgtggcga 184
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 102 tctgtgtggtcctctgtgagcggcgtggaactacagcagagtgacgccttctcctgtggcga 161
QY 185 gtgggggccccttcacagcgcctcatcttctcctgtcagcgcgcagacatccccaatgg 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 162 gtgggggccccttcacagcgcctcatcttctcctgtcagcgcgcagacatccccaatgg 221
QY 245 cttaatagttaatgtaagtcgtgttctcctgtcggggaaccccggaacgcgtgtcgaatgcc 304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 222 cttaacgggctgtctcctcgtgttctctgtagcgaccccgagacacgcgtgtcgggtgcc 281
QY 305 ggaacgcggaaactgaaagcgtgtgacgaacaacagttcccggtgcggtgcgggga 364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 282 ggaacgcggaaactgaaagcgtgtgacgaacaacagttcccggtgcggtgcgggga 341
QY 365 cggccgcgaggtgtgcccaacagctgacgcgcgtacgcgctgcgccacatcgcaactctc 424

```

[illegible]

Db	1422	ggtgggcaagtttgtagtcacgctgccttccatggtctacgtgtctacacgcgagct	1481
Oy	1499	ctaccacaccctgtgcaagaaatgvcggtggggttcacatccacgcctccagaatggg	1558
Db	1482	gtatccacacagtgtagagaacaatgvgtgggagtcacgtccacagatccgcctggg	1541
Oy	1559	cagcatcatgtcccccctactctgttttactcgtggtgtctaaacagaatgcctgcctaat	1618
Db	1542	cagcatctgtctcctcctactctgttttactctgggtgctcagacgcctctcgcctacat	1601
Oy	1619	cgtagtggtgtagtcgactgtctcgtatcgtgaatctccaccccttttttccttgaagt	1678
Db	1602	tctctatgggaagtcctgaccatcctgtacagccatcctcactcctgtttctccagagagct	1661
Oy	1679	gggaatgactctctccagaaaccttagagcagatgcagaaagtgaatggtttcagttcgg	1738
Db	1662	cggtaccacacccacccagacacatctgaccagatgcttaagtagtcaaaagatgaacacag	1721
Oy	1739	gaaaaaacaagaagactcaatgtagagacagaagaatc	1776
Db	1722	aaaactccaagtcaacacagaagtgttaagaatgctc	1759
RESULT 5			
AAS67216/c			
ID	AAS67216 standard; cDNA; 3261 BP.		
XX	AAS67216;		
AC			
DT	13-FEB-2002 (first entry)		
XX			
DE	DNA encoding novel human diagnostic protein #3020.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
XX	23-AUG-2000; 2000US-0649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmanac RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
XX			
DR	P-PSDB; ABG03029.		
XX			
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity -		
XX			
PS	Claim 1, SEQ ID NO 3020; 103bp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		







Oy	1684	tgaactcttcagaaccttagagcagatgcagaagtgaattggttcaagtctgggaaaa	1743
Db	1603	tcctctccaccataccacattgaccagatgctaagggtcaaaagaattaacaattgycaaa	1662
Oy	1744	aacacagagactcaattgatgagacagaana	1771
Db	1663	tccaaagccagacagaagaatgcataaaaaga	1690
RESULT 7			
ID	AAA09890		
XX	AAA09890 standard; cDNA; 1888 BP.		
AC	AAA09890;		
XX			
DT	05-JUL-2000 (first entry)		
XX			
DE	Mouse OCTN2 cDNA sequence.		
XX			
KM	Organic cation transportation; mouse; carnitine transporter protein;		
XX	OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;		
KW	juvenile visceral steatosis; ss.		
XX			
OS	Mus musculus.		
XX			
PN	WO200014210-A1.		
XX			
PD	16-MAR-2000.		
XX			
PF	07-SEP-1999; 99WO-JP04853.		
XX			
PR	07-SEP-1998; 98JP-0252683.		
XX			
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.		
XX			
PI	Nezu J, Oku A;		
XX			
DR	WPI; 2000-256966/22.		
XX			
P-PSDB:	AAV83930.		
PT	Systemic carnitine deficiency gene OCTN2 encoding part of organic		
XX	cation transporter, useful as diagnostic tool -		
XX			
PS	Example 1; Page 51-57; 106pp; Japanese.		
XX			
CC	This sequence represents the cDNA encoding the mouse carnitine		
CC	transporter protein OCTN2. The corresponding human sequence (AAA09889)		
CC	can be used as a target for diagnosis of systemic carnitine deficiency		
CC	by detecting the presence of mutations in the sequence, especially seen		
CC	in the disease juvenile visceral steatosis (jvs). The wild type OCTN2		
XX	gene can be used in the gene therapy of the disease state.		
XX			
SO	Sequence 1888 BP; 392 A; 509 C; 486 G; 501 T; 0 other;		
Query Match 46.6%; Score 994; DB 21; Length 1888;			
Best Local Similarity 75.7%; Pred. No. 1.9e-246;			
Matches 1247; Conservative 0; Mismatches 395; Indels 6; Gaps 1;			
Oy	130	gaagcgacgttggaagaacatgacgagactaagacgaagttagtcgcttcctgagcgagttgg	189
Db	43	gggagagctgaggagcgacatgaggggactacagcagaggtgacgccttcctcagcgagttggg	102
Oy	190	gagcccttcacagcgccctcatcttcttcctgctcacgacgcacatcalcccacattgctcca	249
Db	103	ggcccttcacagcgccctcatcttcttcctgctcacgacgcacatcalcccacattgctcca	162
Oy	250	atggttatgtacagcctggttccttgscgggggagcccccgagacacgcgttgtgafgcccagag	309
Db	163	atggtatgtacacatcggttctctgscgggggagcccccgagacacgcgttgtgcttggctccaca	222
Oy	310	cacgaacctcgaacacgcctctgscgacaaacagttgccgcctgscgtgscgtgscgagcgc	369

Db	223	ccgtcaaacctgagcagcgctgtgctgcaacacagatcccttggttggtagagcaagacgagac	282
Oy	370	gcgaggtgccccacacgctctgacgcgtacacgggtctgcacacacatccgacactctctcgtgcgc	429
Db	283	gacaggtgctctacaagaatgctgcgcgtctacagccttggccccacatctgcacattctctgcgc	342
Oy	430	tcgggtctggagccggggcgcgcgtgagacgttggacgttgggacagcttggagcagagagctgtcctgg	489
Db	343	taaggcttgagacccgggcggagcgttgagaccttgagagcagctgtgagcagagagagagctgtcctgg	402
Oy	490	atggcctggagagttcagccacggagcgccttactcttccacccgtctgttcacccgctgttcacccgggtgaactctgg	549
Db	403	atggcctggagagttcagcacaaggacgcttctctcgttccacacatctgtgacaagagctggagacccgg	462
Oy	550	tgtgtgtaggacacacttggaaagctgtccctccacacacccctgttcttcgtgagccgtgtcc	609
Db	463	tgtgtgtaggagatgagcttggaaagccccatccacacccctgttcttcgtgagctgtgtga	522
Oy	610	tcggctctctctgtctgcgggcagcgtgtcagacaggtttggcagaggttggcaggaagagcttctctcgg	669
Db	523	tgggcctcttcattctcagacagcctctcagacagagtttgcctgcagaaagtgtgtcttct	582
Oy	670	caaacctggtctgacagagcttgcgttcaagttctcctgtgaaagtcttctccacacagctgggagaa	729
Db	583	tgcacatgggcagatgcagactgtgcttaagcttccctgcgaagctctctctgtgaaactctgagaa	642
Oy	730	tgtctacctgtgtattctgtcatcgttggcagatgggcagagcttccacactatgtgtgtagacct	789
Db	643	tgtttcacagtgcttcttctgtccctgtgtggcagatgggtgtagatcccaactcagctgtgcagcgt	702
Oy	790	tcaatactaggaacagaagaatctctgycagatcagttcgtatataatctctacattagagag	849
Db	703	tgtctcctgggaacagaagaatcttctccaaagtcaatctgaattatattctgcacacttagagag	762
Oy	850	tgtgcacacttttctgcagttgtggtctaatgctgtgtctgcacactggttctgttacttccatcaagag	909
Db	763	tttgcataattttaaigtgcttgtgtctatggttgcctgcacactgttgcatacttccatcaagag	822
Oy	910	actggcgagagatgcgcgtgcgtgcgtctacaggtgcgcggagagtgctgtgtctccgcgttggct	969
Db	823	actggcgagagatgcgcgtgcgtgcgtctacactgtgcacaggggtgtcattgttggggcctctgtgt	882
Oy	970	ggttcatctctgaatcctcccgatctgcatactccacagagaagaatttaaggagagctgtgaag	1029
Db	883	ggttcatcctcctggtctcccaacgagtgtctactctctcaaggccgaattaaagagagcagagag	942
Oy	1030	atatatcccaaaaaagctgcgcaaaaaatgacaacacacagctgttaccagcaggtgatatatttgatt	1089
Db	943	tgtatcatctcgcaaaagctgtgcacaaatctcaatctgttctgacacatccatctatcttcgactc	1002
Oy	1090	ctgtgagag-----gagctaatccctctgagcagcagagaagctttcatctctgagactgt	1143
Db	1003	caagtgaggttacaagaacttaaatctctacggaagctccagttgcacacacattatagatctga	1062
Oy	1144	tcaagagactcggaalatgtccataatgacatcatatgtcttctgtcgtatgtagtgcagcct	1203
Db	1063	tcgcgaacacggaatatacagggtctacacacatcatgtctataatccctgttgcgtgacatat	1122
Oy	1204	cagttgggttactctgtcgtctgtctctgtgagatgctccttaattacaatggagagctactactga	1263
Db	1123	cagttgggtctatcttgtgactactctcttgcacactcccttaactgtgcagtgagagacatcatgtga	1182
Oy	1264	actgtttctctctcgtcctgtagtaaatctcaggttcatcatcaacgacctgtgcgtgactatg	1323
Db	1183	actgtctctctactctgtgcgcgtgtgtgaagttccacgactatgtgtcgtgcctgtgtgtgtgc	1242
Oy	1324	gaacgctgcgcccagcggtataatacatcagctcgaagtactgtctcgtggagagaggtgtgcttc	1383
Db	1243	agtaactgcccccgagatatctactctcgcgccttcttcctcgtgtggtgcagttgctcttc	1302
Oy	1384	tcttcatctcaactgtgatcactgtgtgatatatacttctatcatcagctgtgtgtagctgg	1443
Db	1303	tcttcatctgacgctgtgtgtccttccaaatgttgttactgttgcacacgacctgtgtgtgtgtgtg	1362





```
RESULT 10
AAH92241/C
ID AAH92241 standard; DNA; 700 BP.
XX
AC AAH92241;
XX
DT 09-OCT-2001 (first entry)
XX
DE Human inflammatory bowel disease related gene fragment IGR2253a.
XX
KW Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;
single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;
chromosome 5q31-33; forensic test; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200142511-A2.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33632.
XX
PR 10-DEC-1999; 99US-0170257.
XX
PR 10-APR-2000; 2000US-0196046.
XX
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PA (ELI ) ELLIPSIS BIOTHERAPEUTICS CORP.
XX
PI Daly M, Hudson TJ, Lander ES, Rouns J, Siminovitch K;
DR WPI; 2001-367874/38.
XX
PT Testing for the presence of polymorphisms associated with inflammatory
PT bowel disease, using a hybridization assay -
XX
PS Disclosure: Page 159; 463pp; English.
XX
CC The present invention describes a method for detecting the presence of
CC polymorphisms associated with inflammatory bowel diseases such as
CC ulcerative colitis and Crohn's disease. The methods can be used to detect
CC the presence of genetic polymorphisms associated with inflammatory bowel
CC disease and correlating their occurrence with disease states. They may be
CC used in this way for phenotypic correlations, forensics, paternity
CC testing, medicine and genetic analysis. The present sequence is a gene
CC containing a polymorphic site described in the exemplification of the
CC invention.
XX
SQ Sequence 700 BP; 114 A; 232 C; 260 G; 94 T; 0 other;

Query Match 17.0%; Score 363.8; DB 22; Length 700;
Best Local Similarity 92.3%; Pred. No. 8.4e-84;
Matches 383; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 125 ttctgagcgagcagtcggaagcatgaggaactgaagagtgatgccttcctggcgca 184
DB 552 TCTGTGGGCTCTGAGGCGGCGATCGGGACTACGACAGCAGCTGACCGCTTCTGCGCGA 493
QY 185 gtgggggccccttcagcgccatctcttcctgtctgtagcgagcatcatccccaatgg 244
DB 492 GTGGGGGCCCCCTTTCAGCGCCTCATCTTCTCTGCTCAGCGCGACATCATCCCAATGG 433
QY 245 ctccaatggtatgtaagtcgtctctgtgagcgagcccgagagcagctgtcgatgac 304
DB 432 CTTCACGGCGCTGCTCTCGTCTCTGATAGCGACCCCGGAGCAGCGCTCCGGGTCC 373
QY 305 ggaagcgcgaaacctgaaacagcgctgtgagcaacaacatgtcccgctgaggtcgagga 364
DB 372 GGACGCGCGAAGCTGACAGCGCTGCGCAACACACTCTCCACATGCGGCTGCGGGA 313
QY 365 cggcgcgagtggtcccaacagctgacagcgctacacggctgcgaacaactcttc 424
DB 312 CGGCGGCGAGGTGCCCAACAGCTGCCGCGCTACCGGCTCCGCCACCACTGCTTC 253
```

```
QY 425 ggcgcctggagctgagccggggcgagcagctgagcctggggcagcctgagcagagagctg 484
DB 252 GGGCGCTCGGCTGAGCGCGCGCGACGTGAGACTGTGGGCGACCTGGAGCAGAGAGCTG 193
QY 485 cctgagtgagctggaggaattcagacagcgtctcctctccacgctgtgagcag 539
DB 192 TCTGATGGCTGGGAGTTCAGTCAAGACGTCTACTGTCTCCACATTTGTGACCGAG 138

RESULT 11
AAA09888
ID AAA09888 standard; DNA; 25871 BP.
XX
AC AAA09888;
XX
DT 05-JUL-2000 (first entry)
XX
DE Human genomic OCTN2 sequence.
XX
KW Organic cation transportation; human; carnitine transporter protein;
KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
KW juvenile visceral steatosis; ds.
XX
OS Homo sapiens.
XX
PN WO200014210-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-JP04853.
XX
PR 07-SEP-1998; 98JP-0252683.
XX
PA (CHUG ) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Nezu J, Oku A;
DR WPI; 2000-256966/22.
XX
PT Systemic carnitine deficiency gene OCTN2 encoding part of organic
PT cation transporter, useful as diagnostic tool -
XX
PS Claim 1; Page 60-93; 106pp; Japanese.
XX
CC This sequence represents the genomic DNA encoding the human carnitine
CC transporter protein OCTN2 (A183929). The complete genomic sequence
CC contains 9 introns dispersed throughout the gene. The gene can be used
CC as a target for diagnosis of systemic carnitine deficiency by detecting
CC the presence of mutations in the sequence, especially seen in the
CC disease juvenile visceral steatosis (jvs). The wild type OCTN2 gene can
CC be used in the gene therapy of the disease state.
XX
SQ Sequence 25871 BP; 6356 A; 6094 C; 6408 G; 7013 T; 0 other;

Query Match 17.0%; Score 363.8; DB 21; Length 25871;
Best Local Similarity 92.3%; Pred. No. 5.7e-83;
Matches 383; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 125 ttctgagcgagctggaagcatgcggaactagagagtgatgccttcctggcgca 184
DB 200 tctgtgagccttcgagggcgagctagcgactagagagtgatgacgccttcctggcgca 259
QY 185 gtgggggccccttcagcgccatctcttcctgtctcagcgagcagatcatccccaatgg 244
DB 260 gtgggggccccttcagcgccatctcttcctgtctcagcgagcagatcatccccaatgg 319
QY 245 ctccaatggtatgtaagtcgtctctgtgagcgagcccgagagcagctgtcgatgac 304
DB 320 ctccaacggagctgtccctcgltctctgtagagagaccggagcagcgtgtccgggtgac 379
QY 305 ggaagcgcgaaacctgaaacagcgctgtgagcaacaacatgttcgcgtgaggtcgagga 364
```

Db 380 ggaacgcgcgaactgagcagcctgtagcaacacacatgtcccaactgagctgcgagga 439  
Oy 365 ggcgcgcgaggtgcccacagctgcaacgcgttaccgctgcacacatctgcaattctc 424  
Db 440 cgcgcgcgaggtgcccacagctgcccgcctaccgctgcacacatctgcaattctc 499  
Oy 425 ggcgcgcgaggtgcccacagctgcccgcctaccgctgcacacatctgcaattctc 484  
Db 500 ggcgcgcgaggtgcccacagctgcccgcctaccgctgcacacatctgcaattctc 559  
Oy 485 cctgagatgctgagagtcacgacgagcgtctacccttcacacgctgtagccgag 539  
Db 560 tctgagatgctgagagtcacgacgagcgtctacccttcacacatctgtagccgag 614

RESULT 12  
AAH92293/C  
ID AAH92293 standard; DNA: 700 BP.  
XX  
AC AAH92293;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Human inflammatory bowel disease related gene fragment IGR2305a.  
XX  
KW Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
KW single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;  
KW chromosome 5q31-33; forensic test; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200142511-A2.  
XX PD 14-JUN-2001.  
XX PF 11-DEC-2000; 2000WO-US33632.  
XX PR 10-DEC-1999; 99US-0170257.  
XX PR 10-APR-2000; 2000US-0196046.  
XX PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
XX PA (ELLIPIS ) ELLIPIS BIOTHERAPEUTICS CORP.  
XX PI Daly M, Hudson TJ, Lander ES, Rioux J, Siminovitch K;  
XX WPI: 2001-367874/38.  
XX DR  
XX PT Testing for the presence of polymorphisms associated with inflammatory  
XX PT bowel disease, using a hybridization assay -  
XX PS  
XX PS Disclosure; Page 174-175; 463pp; English.  
XX  
CC The present invention describes a method for detecting the presence of  
CC polymorphisms associated with inflammatory bowel diseases such as  
CC ulcerative colitis and Crohn's disease. The methods can be used to detect  
CC the presence of genetic polymorphisms associated with inflammatory bowel  
CC disease and correlating their occurrence with disease states. They may be  
CC used in this way for phenotypic correlations, forensics, paternity  
CC testing, medicine and genetic analysis. The present sequence is a gene  
CC containing a polymorphic site described in the exemplification of the  
CC invention.  
XX  
SQ Sequence 700 BP; 197 A; 132 C; 139 G; 232 T; 0 other;

Query Match 14.7%; Score 314; DB 22; Length 700;  
Best Local Similarity 100.0%; Pred. No. 6; 2e-71;  
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1822 ggttaagtgaataaataagacccctgtgagaataatctgttccactgaat 1881  
Db 700 GGTAAAGTGAATAAATAAGACCCCTGTGAGAAATCTGTTCCTCCACTGAAT 641

Oy 1882 ggaactgatacgaattagacccaataatgaacctgtgatacaagaataatgctcgtatrac 1941  
Db 640 GGACTACTGTATACGATTACACCAATAATGAACCTTGCTATCAAGAAATGCTGTCATAC 581  
Oy 1942 agtaactctgataatgattcttcagaataatgctcgttcaacaacacattctag 2001  
Db 580 AGTAAACTCTGATATGATTCTTCAGATATGTCCTTGCTTACAAACCAACATTTCAG 521  
Oy 2002 agagcttccttaccatcaatcaatgaatgattgtaagaatgcttgaacaatgtt 2061  
Db 520 AGAGTCTCCTTACTCATTTAATTCAAATGAATGATGATGATGATGATGATGATGATGAT 461  
Oy 2062 agtcaagagctgtaataatatacaataaagattacacattccatcaatcaataat 2121  
Db 460 AGTCAGAGACTGCTAATAATACATATTAAGATTACACTCTATTCATCAATCAATAACT 401  
Oy 2122 atccaataataaat 2135  
Db 400 ATCCAATAATAAAT 387

RESULT 13  
AAH92294/C  
ID AAH92294 standard; DNA: 700 BP.  
XX  
AC AAH92294;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Human inflammatory bowel disease related gene fragment IGR2306a.  
XX  
KW Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
KW single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;  
KW chromosome 5q31-33; forensic test; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200142511-A2.  
XX PD 14-JUN-2001.  
XX PF 11-DEC-2000; 2000WO-US33632.  
XX PR 10-DEC-1999; 99US-0170257.  
XX PR 10-APR-2000; 2000US-0196046.  
XX PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
XX PA (ELLIPIS ) ELLIPIS BIOTHERAPEUTICS CORP.  
XX PI Daly M, Hudson TJ, Lander ES, Rioux J, Siminovitch K;  
XX WPI: 2001-367874/38.  
XX DR  
XX PT Testing for the presence of polymorphisms associated with inflammatory  
XX PT bowel disease, using a hybridization assay -  
XX PS  
XX PS Disclosure; Page 175; 463pp; English.  
XX  
CC The present invention describes a method for detecting the presence of  
CC polymorphisms associated with inflammatory bowel diseases such as  
CC ulcerative colitis and Crohn's disease. The methods can be used to detect  
CC the presence of genetic polymorphisms associated with inflammatory bowel  
CC disease and correlating their occurrence with disease states. They may be  
CC used in this way for phenotypic correlations, forensics, paternity  
CC testing, medicine and genetic analysis. The present sequence is a gene  
CC containing a polymorphic site described in the exemplification of the  
CC invention.  
XX  
SQ Sequence 700 BP; 224 A; 120 C; 124 G; 232 T; 0 other;

Query Match 13.9%; Score 297; DB 22; Length 700;

Best Local Similarity 100.0%; Pred. No. 1,5e-66;  
Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1726 ggtcaagatctggaagaaacagaagactcaatggaagacagaagaaatcccaaggttc 1785  
|||||  
Db 297 GGTTCAGATCTGGGAAAAAACAGAGACTCAATGAGACAGAAATAATCCCAAGGTTTC 238  
|||||

QY 1786 taatactcattctgaaataatctaccatttggtgaagtaaaaaagaaaaata 1845  
|||||  
Db 237 TAATACATCTGATTTGMAAAATATCTACCCATTGGTGAAGTAAACAGAAAAATA 178  
|||||

QY 1846 agaccctgtagaagaaatcgtgtcccactgaaatggactgtaactgtaacgattgacac 1905  
|||||  
Db 177 AGACCCCTGGGAGAAATTCGTTGCCACGTAATGGACGACTGTAACGATTGACACCC 118  
|||||

QY 1906 aaaaagaacctgtctatacagaagaatgctcgtatatacagtaaaacttgatgattctcca 1965  
|||||  
Db 117 AAAATGAACCTTGCTATAGAAATGCTGCTATACAGTAAACCTGGATGATTCTTCCA 58  
|||||

QY 1966 gataatgctcctgtcttcaaacaccattcttagagagtccttactaatat 2022  
|||||  
Db 57 GATAATGTCCTTGCTTTACAAACCAACATTCTTAGAGAGTCTCCTTACTATTAT 1

RESULT 14  
AAH92756/c  
ID AAH92756 standard; DNA: 700 BP.  
XX  
AC AAH92756;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Human inflammatory bowel disease related gene fragment IGR3081a.  
XX  
KW Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
KM single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;  
KW chromosome 5q31-33; forensic test; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200142511-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 11-DEC-2000; 2000WO-US33632.  
XX  
PR 10-DEC-1999; 99US-0170257.  
PR 10-APR-2000; 2000US-0196046.  
XX  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PA (ELLI-) ELLIPSIS BIOTHERAPEUTICS CORP.  
XX  
PI Daly M, Hudson TJ, Lander ES, Rioux J, Siminovitch K;  
XX  
DR WPI: 2001-367874/38.  
XX  
PT Testing for the presence of polymorphisms associated with inflammatory  
XX bowel disease, using a hybridization assay -  
XX  
PS Disclosure: Page 305; 463pp; English.  
XX  
CC The present invention describes a method for detecting the presence of  
CC polymorphisms associated with inflammatory bowel diseases such as  
CC ulcerative colitis and Crohn's disease. The methods can be used to detect  
CC the presence of genetic polymorphisms associated with inflammatory bowel  
CC disease and correlating their occurrence with disease states. They may be  
CC used in this way for phenotypic correlations, forensics, paternity  
CC testing, medicine and genetic analysis. The present sequence is a gene  
CC containing a polymorphic site described in the exemplification of the  
CC invention.  
XX  
SQ Sequence 700 BP; 134 A; 219 C; 217 G; 130 T; 0 other;

Query Match 11.0%; Score 234; DB 22; Length 700;  
Best Local Similarity 100.0%; Pred. No. 2.9e-50;  
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 gaccgcgcgaacctgagcaagcctggtcgcaacaacagttcccgctggtcggtcggtgagac 365  
|||||  
Db 700 GAGCCCGCGAAGCTGAGACAGCGCTGGCCGACAGAGTGTCCTCCGGTGGCGGTGGCGGAG 641  
|||||

QY 366 ggcgcgaggtgtcccaacagctgagcgcgtaccggtctgcacatcgccaattctcg 425  
|||||  
Db 640 GGCCTGCGAGGTGCCCGCACACTGACGCGCTACCGCTGCCACATCGCCAACTTCTCG 581  
|||||

QY 426 ggcctcgagctgagccggggcgagcgtgagcttgggagctggagcagagagctgc 485  
|||||  
Db 580 GCGCTGCGGCTGAGACCGCGGCGCGAGCTGAGCTGGGGCAGCTGAGCAGGAGAGACTGC 521  
|||||

QY 486 ctggaatgctggtgagttcagccagagcgtctacctgtccacgctggtgaccgag 539  
|||||  
Db 520 CTGATGCGTGGGAGTTGACCCAGAGAGCTGTACTGTCCACCGTGTGACCGAG 467  
|||||

RESULT 15  
AAH92677/c  
ID AAH92677 standard; DNA: 700 BP.  
XX  
AC AAH92677;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Human inflammatory bowel disease related gene fragment IGR3002a.  
XX  
KW Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
KM single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;  
KW chromosome 5q31-33; forensic test; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200142511-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 11-DEC-2000; 2000WO-US33632.  
XX  
PR 10-DEC-1999; 99US-0170257.  
PR 10-APR-2000; 2000US-0196046.  
XX  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PA (ELLI-) ELLIPSIS BIOTHERAPEUTICS CORP.  
XX  
PI Daly M, Hudson TJ, Lander ES, Rioux J, Siminovitch K;  
XX  
DR WPI: 2001-367874/38.  
XX  
PT Testing for the presence of polymorphisms associated with inflammatory  
XX bowel disease, using a hybridization assay -  
XX  
PS Disclosure: Page 282; 463pp; English.  
XX  
CC The present invention describes a method for detecting the presence of  
CC polymorphisms associated with inflammatory bowel diseases such as  
CC ulcerative colitis and Crohn's disease. The methods can be used to detect  
CC the presence of genetic polymorphisms associated with inflammatory bowel  
CC disease and correlating their occurrence with disease states. They may be  
CC used in this way for phenotypic correlations, forensics, paternity  
CC testing, medicine and genetic analysis. The present sequence is a gene  
CC containing a polymorphic site described in the exemplification of the  
CC invention.  
XX  
SQ Sequence 700 BP; 229 A; 139 C; 129 G; 197 T; 6 other;

Query Match 10.2%; Score 217.6; DB 22; Length 700;



Best Local Similarity	91.6%;	Pred. No. 5.1e-46;
Matches	229; Conservative	0; Mismatches 21; Indels 0; Gaps 0;

Matches 229; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

1168 tgaccattatgtcttctgtctatgatgtctgacctcagtggttactttgctctctc 1227

Db 308 TGTCA TTTTACCTTCTTCTTTCAGGATGCTGACCTCAGTGGTTACTTTGCTCTGCTC 249

QY 1228 tggatgctcctaattacatgagatgcctaactgttccctctgcctgatg 1287

Db 248 TGGATGCTCCTAATTACATGGAGATGCCCTACCTGAACGTGTTCCCTCTCTGCCCTTGATG 189

QY 1288 aaattcagcttacatcagcctgctgctattgcgaacgctgccaggcgttatatca 1347

Db 188 AAATTCCAGCTTACATTACAGCCTGGCTGCTATTGGGAACCCCTGCCCAGGCGTATATCA 129

QY 1348 tagctgcagtactgttctgggagagagtgtgcttctcttcaactgttacctgtg 1407

Db 128 TAGCTGCAGTACTGTCTCTGGGAGGAGGTGTGCTTCTCTTCATTCAACTGGTACCTGTGG 69

QY	1408	attattactt	1417
----	------	------------	------

Db 68 GTANNAAGTT 59

Search completed: July 17, 2002, 01:51:58  
Job time: 18884 sec

Job time: 18884 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 16, 2002, 20:07:29 ; Search time 135.85 seconds  
(without alignments)  
3860.343 Million cell updates/sec

Title: US-09-521-195B-2  
Perfect score: 2135  
Sequence: 1 cccgcgtcgcgcgcccaat.....aatactatccaataaaat 2135

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.2	9.1	1896	US-08-501-572-6	Sequence 6, Appl
2	195.2	9.1	1896	US-09-040-444-6	Sequence 6, Appl
3	175	8.2	1882	US-08-501-572-4	Sequence 4, Appl
4	175	8.2	1882	US-09-040-444-4	Sequence 4, Appl
5	174.8	8.2	1885	US-08-501-572-5	Sequence 5, Appl
6	174.8	8.2	1885	US-09-040-444-5	Sequence 5, Appl
7	107	5.0	2102	US-08-647-397-1	Sequence 1, Appl
8	104	4.9	370	US-08-592-126-108	Sequence 108, App
9	104	4.9	481	US-08-592-126-109	Sequence 109, App
10	55	2.6	1794	PCT-US95-14418-3	Sequence 3, Appl
11	55	2.6	1794	PCT-US95-15327-3	Sequence 3, Appl
12	55	2.6	3048	PCT-US95-14418-1	Sequence 1, Appl
13	55	2.6	3048	PCT-US95-15327-1	Sequence 1, Appl
14	53.4	2.5	1686	US-08-648-657-14	Sequence 14, Appl
15	53.4	2.5	2504	US-08-073-384C-3	Sequence 3, Appl
16	53.4	2.5	2504	US-08-254-359A-3	Sequence 3, Appl
17	53.4	2.5	2504	US-08-483-043-3	Sequence 3, Appl
18	53.4	2.5	2504	US-08-481-238-3	Sequence 3, Appl
19	53.4	2.5	2504	US-08-471-066B-3	Sequence 3, Appl
20	53.4	2.5	2504	US-08-484-956-3	Sequence 3, Appl
21	53.4	2.5	2504	US-08-757-653-3	Sequence 3, Appl
22	53.4	2.5	2504	US-08-599-491-3	Sequence 3, Appl
23	53.4	2.5	2504	US-08-756-386-3	Sequence 3, Appl
24	53.4	2.5	2504	US-08-823-516-3	Sequence 3, Appl
25	53.4	2.5	2504	US-08-682-853A-3	Sequence 3, Appl
26	53.4	2.5	2504	US-08-759-038-3	Sequence 3, Appl
27	53.4	2.5	2504	US-08-758-314-3	Sequence 3, Appl

28	53.4	2.5	2504	US-09-350-309-3	Sequence 3, Appl
29	53.4	2.5	2505	US-07-977-434-9	Sequence 9, Appl
30	53.4	2.5	2505	US-08-458-819-9	Sequence 9, Appl
31	53.4	2.5	2505	PCT-US91-07035-9	Sequence 9, Appl
32	53.4	2.5	2640	US-08-384-480-30	Sequence 30, Appl
33	53.4	2.5	2640	US-08-459-383-30	Sequence 30, Appl
34	50.4	2.4	7218	US-08-232-463-14	Sequence 14, Appl
35	50	2.3	7898	US-08-984-709A-49	Sequence 49, Appl
36	48.4	2.3	1490	US-08-964-127-3	Sequence 3, Appl
37	48.4	2.3	1490	US-09-496-692-3	Sequence 3, Appl
38	48.4	2.3	2460	US-08-964-127-1	Sequence 1, Appl
39	48.4	2.3	2460	US-09-496-682-1	Sequence 1, Appl
40	48.2	2.3	2502	US-08-073-384C-7	Sequence 7, Appl
41	48.2	2.3	2502	US-08-254-359A-7	Sequence 7, Appl
42	48.2	2.3	2502	US-08-483-043-7	Sequence 7, Appl
43	48.2	2.3	2502	US-08-481-238-7	Sequence 7, Appl
44	48.2	2.3	2502	US-08-471-066B-7	Sequence 7, Appl
45	48.2	2.3	2502	US-08-484-956-7	Sequence 7, Appl

## ALIGNMENTS

```
RESULT 1
US-08-501-572-6
; Sequence 6, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport of protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farbow,Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4000
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-501-572-6
```

Query Match 9.1%; Score 195.2; DB 3; Length 1896;  
Best Local Similarity 48.5%; Pred. No. 5,6e-42;  
Matches 635; Conservative 0; Mismatches 663; Indels 12; Gaps 3;

QY 440 gccggggcgacgtgacctggcgagctggagcagagagactgcctgagctgagga 499

```
Db 465 GGCACACCTGGAGACACCAAGAGCCGCTGCGACCTGGCCCCCTCCGGAGCGCTGGGT 524
Qy 500 gtccagccagagacgtctacactgtccacgctgtgacccagtgagtaactgtgtgtgagga 559
Db 525 GTACGAGAGCGCTGGC-----TCGTCCATCTCTCACCGAGTTTAACGTGTATGTGCCAA 578
Qy 560 caactggaaagtgtccctaccacactccctgtctctgtagcgtgtccctccgtccct 619
Db 579 CTCCTGTGATGTGGACCTATTCACGATTCAGTGAATGAGATTCCTTATTTGGCTCAT 638
Qy 620 cgtgtccggcagcagctgtcagacaggtttggcagaagaagcttccctccgaacctg 679
Db 639 GAGTATCGGCTCATAGACGACGAGGTTGGCCGTTAAGCTTCGCTCCACACTGAGTCT 698
Qy 680 tgcagacgtgtccacgtctccacagtccacagtccacagtccacagtccacagtccac 739
Db 699 CATTAATGCTGACGCTGAGCTTCATGCTCATGCTTCCTCCCACTATACGTGATTTAT 758
Qy 740 gtatttgcacgtgtgagcagatggcagacatccacactatgtgtgagctccatacag 799
Db 759 TTTTGGCTTAATCCAAAGACTGTGTCAGCAAGCAGCTGTTAATAGGCTCATGCTCAT 818
Qy 800 aacagaaatcttggcagagtcagtcgtatatactctacataagagctgtcgaact 859
Db 819 TACAGAAATTTGTTGGGGGAGATATCGAAGAACAGTGGGATTTTTCACAAATTGC--- 875
Qy 860 ttctgcagttgtcgtatgtcgtccacgtgttgccttaccacagacagacagcgt 919
Db 876 CTATACAGTTGGGCTCCGCTGCTGCTAGCTGGGGTGGCTTACGCACTCTCTCAGAGGT 935
Qy 920 gctgtcgtgtgagcgtgtgagcgtgtgagcgtgtgtgtgtgtgtgtgtgtgtgtgt 979
Db 936 GTTGCAGTTCACAGTGTGCTGCTGCCCACTTCTTCTTGTGCTCATATGCTGCTGCTAC 995
Qy 980 tgaactcccccagtgctgatatcccaagaagatttgagaagctgtgaagatatacaca 1039
Db 996 TGAAGTCTCCAGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
Qy 1040 aaaaagctgcacaaatgacacacacagctgtacacagctgtatattgtgtgtgtgt 1099
Db 1056 GCACATCGCAAAAGAAATGGAATCTCTACCGCCCTCCCTTCAGCCCTGAGAACTTGA 1115
Qy 1100 gctaaatccctgtgaagcagcagaagcttctacatcgtgacgtgtcagagctcgaa 1159
Db 1116 AGAGAAACTGGCAGAAATGGAATGGAATCTCTACCGCCCTCCCTGAGAACTCTCAGAT 1175
Qy 1160 tgcataatgacacattatgtcttgcgtatagatgtgtgacccagtggtgtgtgtgt 1219
Db 1176 AAGGAAACATATATGATATGATGTATGATGATGATGATGATGATGATGATGATGATG 1235
Qy 1220 tctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1279
Db 1236 CCTCATCATGCAATGGGCTTTGCAAGTGACATATATCACTGATTTCTTCTCATCTGTC 1295
Qy 1280 ctgtatgaaatccaggttccatcaacagcgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1339
Db 1296 CCGGTGTAATTCAGCTGCTTCCTTATGATGATGATGATGATGATGATGATGATGATGAT 1355
Qy 1340 ttatacatagctgcagacgtgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1399
Db 1356 TTACCTTGGGCTGCTCATCAATATGTTGCCAGGGGACCGCTGTGCTGCTGCTGCTGCT 1415
Qy 1400 acctgtgagatattactctatccatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1459
Db 1416 ACCTGTGATCTACAAATGAGTAAATATATATCTCATGCTGGGAGAGATGGGAGATCAC 1475
Qy 1460 ctctgtcttcccaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1519
Db 1476 AATGGCCCTANGAGATAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1535
Qy 1520 catgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1579

Db 1536 TCTTGGGCTCCACATCTGTTCTCATATGTGTGACATTTGTGGCATCATACGCCATTTCT 1595
Qy 1580 tgttacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1636
Db 1596 GGTCTACCGGCTCATCAATCTGCTGTGAGCTCCCGCTGATGTTTTCGGCTACTTGG 1655
Qy 1637 tgtctgtatggaatcttccaccccttttccctgaaagtttggaaatgtgacttccaga 1696
Db 1656 CTGTGTTGCTGTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1715
Qy 1697 aaccttagacagatgcagaaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1746
Db 1716 GACCATGAGAGAGCCGAAATATGCAAAAGCAACGAAATAAAGAA 1765

RESULT 2
US-09-040-444-6
; Sequence 6, Application US/09040444
; Patent No. 6063766
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; TITLE OF INVENTION: Transport protein which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/Or Pharmaceuticals,
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner, L.L.P.
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,444
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-040-444-6

Query Match 9.1%; Score 195.2; DB 3; Length 1896;
Best Local Similarity 48.5%; Pred. No. 5,6e-42;
Matches 635; Conservative 0; Mismatches 663; Indels 12; Gaps 3;

Qy 440 gccggggcgagcgtgtgacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 499
Db 465 GGCACACCTGGAGACACCAAGAGCCGCTGCGACCTGGCCCCCTCCGGAGCGCTGGGT 524
Qy 500 gtccagccagagcgtctacactgtccacgctgtgacccagtgagtaactgtgtgtgagga 559
Db 525 GTACGAGAGCGCTGGC-----TCGTCCATCTCTCACCGAGTTTAACGTGTATGTGCCAA 578
Qy 560 caactggaaagtgtccctaccacactccctgtctctgtagcgtgtccctccgtccct 619
```

```

Db      579 CTCTGGAGTGTGGACCTATTCACATCATGATGATGAGATTCCTTTATTTGGCTCTAT 638
Qy      620 cgtgcaggagcgtgtcagacagaatttggcagaagaagctctcttcgcaaccagcgc 679
Db      639 GAGTATGGCTACATACAGACAGAGTTGGCCGTAAGCTTGCTCTTAATCTACAGTCTT 658
Qy      680 tglacagactgttcagctcctcagcagatttctccatcagctgtggagagtgtcactgt 739
Db      699 CATAAATGCTCAGCTGGAGTTCTCATGGCCATTTCCCAACCTATACGAGATGTAAAT 758
Qy      740 gttatttgatcgtgtggcagatggcagacatccactatgtgttagcttcatactagg 799
Db      759 TTTTCGCTTAATCCAGAGACTGGTCACAGCAAGCGGTGGTTAATAGGCTACATCCGAT 818
Qy      800 aacgaatcttctgcaagtcagtcagtcattatattctccatcagctaggagtgagacat 859
Db      819 TACAGAAATTTTGGGGGAGATATCGGAGACAGTGGGATTTTTCACCAAGTTGC--- 875
Qy      860 ttctgcaactgtctatactgtcgtccagctgttcttacttccacagagactgtgcgat 919
Db      876 CTATACAGATTGGGCTCGTGGTGAAGCTGGGGGTGGCTTAGCAGACTTCTCAGCTGAGAGTG 935
Qy      920 gctgctgtgctgagcgtgacggtgtccggagagtgctgtgtcccgctgtgtgttccatcc 979
Db      936 GTTCAGATTACAGTGTGCTGTGCCCAACTTCTTCTCTTATTAATGAGTGAGCATACC 995
Qy      980 tgaatctcccgatgctgatatcccgagagaaatttagagagctgtgaatcaccac 1039
Db      996 TGAGCTCTCCAGGTGGCTGATCTCCAGAAATACAGAAATGCTGAACCATGAGAACTATTA 1055
Qy      1040 aaaaagctgcaaaaatgacaacacacagctgtacacagcagtgatattgtatctgtggagga 1099
Db      1056 GCACATGSCAAGAAAAATGSAAAATCTCTACCGGCTCCCTTAGGGCCGTGACACTTGA 1115
Qy      1100 gctaaatccctgagcagaagaagcttcaatctgtgacctgttcagagctcggaatc 1159
Db      1116 AGAGAAACTGGCAAGAAATGAAACCTTCAATTTCTTGACTTGGTCGAACTCCCTCAGAT 1175
Qy      1160 tgcataatgacattatgtcttctgagctatgagtgacccagtgagtgactcttcgc 1219
Db      1176 AAGGAACATATCTATGATATTTGATGATACACTGTTACAGAGCTGTGCTCTACCAAGG 1235
Qy      1220 tctgtctgtgagtctccataattacatgagatgacgtaccgtgacactgttccctctgc 1279
Db      1236 CCGTATCATGACATAGGGCTTTGCAAGTGACAAATATCTACCTGATTTCTTCTACTCTGTC 1295
Qy      1280 ctgtatgaaatccagcttaccatcagcgtgtgctgtatctgcgaacgtgtgccagcg 1339
Db      1296 CCGGTGTAATTCACAGCTGCTTCATGATCATCTCATATTCGACCGCATCGAGCGCG 1355
Qy      1340 ttatacatagctgagacagctgtctgtgggagaggtgtgtcttctcattcaactggt 1399
Db      1356 TTACCTTTGGGCTGATATAAATATGTTGCAAGGGCAACCTGCTCGGCTCAGTTTTTAT 1415
Qy      1400 acctgtgattattactcttactcattcagctgtgtgtcagctgtgggaaatttggagtaoc 1459
Db      1416 ACGTGTATCTACAAATGGCTAAATAATATATATCTCATGCTTGGGAAATAGGGGATAC 1475
Qy      1460 ctctgtcttccatcagctgtgtatgtcttcaactgtgtgagcttcaaccaacctgtgtcagaa 1519
Db      1476 AATGAGCTATGAGATAGTCTGCTGCTCAATGCTGAGCTGATCCCAATTCATATAGGAA 1535
Qy      1520 catggcgtgtggggtcacatcagcaggtccagagtgaggagcaatcattgcccctactt 1579
Db      1536 TCTTGCGCGCAACATCTGTTCTCTCAATGTGTGACATGTGGGCAATCAGCACTTCTT 1595
Qy      1580 tgttaacctgtgtcttaacaagaat---gttgcctacatcgtcatgggttagtctgac 1636
Db      1596 GGTGTACACGGGCTCAATACATCTGGCTTGAGTCCCGGTGATGTTTCCGGGTACTGG 1655
Qy      1637 tgtcctgattggaatcttcaacccttlttccctgtgaaagtttgggaatgactcttcaga 1696

```

```

Db      1656 CTTCGTTCTGAGAGTCTGCTGCTGTCTCCAGAAACTAAAGGAAAGCTTGGCTGA 1715
Qy      1697 aaccttagcagcagatgcagaagaatgaaatgtgtcagatcctgtggaaaaaaa 1746
Db      1716 GACCATCGAGAGAACCGGAAATATGTGAAGACCAAGAAAAAATAAAGAAA 1765

RESULT 3
US-08-501-572-4
; Sequence 4, Application US/08501572
; Patent No. 6063623
;
GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorbulev, Valentin
; TITLE OF INVENTION: Transport of protein Which effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Finegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/501,572
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Toohey, Kimberlin M
; REGISTRATION NUMBER: 35,391
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-501-572-4

Query Match      8.2%; Score 175; DB 3; Length 1882;
Best Local Similarity 47.5%; Pred. No. 1,3e-36;
Matches 585; Conservative 0; Mismatches 640; Indels 6; Gaps 2;

```

```

Db      674  GTCAGCAAGGCGAGCTGGGCTGCCGCTATACCTTGATCAGACAGATTGTGCGC---TCT 730
QY      822  gtctgattatattctctacatagaggtgacacattttgcagtggtgcatactgtg 881
Db      731  GGCCTACAGAGAAACAGCGGCAATTTGTATACAGATGGCCCTTCACAGTGGGCTACTGGGG 790
QY      882  ctgcacatgttcttactactacacagacatgcagatgcgtctgctgcgcgtgacgtg 941
Db      791  CTTGGCGGGGTGGCTATGCCATTCACAGACTGGCCCTGGCTCCAGCTAGCTGTGTCCTG 850
QY      942  ccggagagtgctgtgtcccgctgtgtgtgtcattctcctgaatctcccgatgctgata 1001
Db      851  CCTACCTTCCTCTCTCTGCTGTATTACGTGTTGTGTCAGAAATCCCGCGGGGCTGTG 910
QY      1002  tcccagagaaattagagagagctgaagatatcatccaaaagcgaagaatgaacac 1061
Db      911  TCCCAAGAAAGAACACACCGGAGCTGTCAAGATATGAGCAATATGCACAGAAAGAGGG 970
QY      1062  acagctgtacacagcagtgatattgtattctgtgagagagtaaatccctgaagcag 1121
Db      971  AAGGTGCTCCTGCTGACCTGAAGATGCTCTGCCCTTGAGAGAGATGCTCAGAAAAGGA 1030
QY      1122  aaagcttcaatcttgacctgttccagacctggaaatattgcataatgacattatgtct 1181
Db      1031  ACTGCTTGTGTTGCCGACCTGTCCGCACTCCCAACCTGAGAGAACACCGTCAATCCTG 1090
QY      1182  ttgcctcatgtagtctgacacctcagctcaggttactctgtctctctctgattgaaatccagcttac 1241
Db      1091  ATGTATCTATGCTGTCTCTTGTGCTGTGCTGACGACAGGGCTTCATCATCATCAGCTGGAGCC 1150
QY      1242  ttacatgtagatgacctgaactgtttccctctctctgattgaaatccagcttac 1301
Db      1151  ACAGGGGCCAACCTCTAGCTGAGACTTCTTTATCTCTCTGAGTGAATCCCGCGGCGC 1210
QY      1302  attacagcctgctgtctatttggaacgctgcgcagggcttatataatagctgcagtactg 1361
Db      1211  TTCAATCATCTCTGCTACCATATGACCGCATATGGCCGATCTACCCAAATGCGGCTCGAAT 1270
QY      1362  ttctgaggagaggtgtgtctctctctcaatcaatgacctgagctgattactactta 1421
Db      1271  CTGGTGAAGGGGGGAGCGCTGCCTCTCATGATCTTTATCCGCAATGAGCTGCATCTGTG 1330
QY      1422  tccattgtctgtgtaactctgaggaaatttggatcacctctgcttctccatgctgtat 1481
Db      1331  AACGTTACCTCTGCTGCTGCTGCGCTGATGAGGGGCCACCATTTGCTCAAGATGATGCTGC 1390
QY      1482  gtcttcactgtcgaactctaacccaaacctgtgcaggaacaatgagcggtggtggtcaatcc 1541
Db      1391  CTGTTGAACGCTGAGCTGATCCCTATCATTCATCAGAAATCTTGGGATGATGATGATGCTCT 1450
QY      1542  acgagctcagagtgtagacatcatctgccccctactctgtttac---ctcggtgttac 1598
Db      1451  GCCCTGTGAGACCTGGGGGATCTTTCACCCCTTCATGAGTTCAGAGCTATATGAAAGTT 1510
QY      1599  aacagaaatgctgcacctacaatcgtcatggtgtgactgactgctcgtatggaatcccaac 1658
Db      1511  TGGCAAGCCCTGCCCTCATTTTGTGTTGTTGGGCTTGTGGCTGACCTGCTGCGGCGCATGACT 1570
QY      1659  ctttttccctgaagaagtttggaatgactcttcacagaacacttagagcagatgcaaaa 1718
Db      1571  CTTCTTCTCCAGAGACCAAGGGTGTGGCTTGTGCTGAGACTTATTTGAAGAGCAGAGAAC 1630
QY      1719  gtgaaatggttcaagatctgaggaaaaaaacaa 1749
Db      1631  CTGGGAGAGAGAAATCAAGGCCAAAGAA 1661

```

RESULT 4  
 US-09-040-444-4  
 ; Sequence 4, Application US/09040444  
 ; Patent No. 6063766  
 ; GENERAL INFORMATION:

```

?      APPLICANT: Koepsell, Hermann
?      APPLICANT: Grundeman, Dirk
?      APPLICANT: Gorboulev, Valentin
?      TITLE OF INVENTION: Transport protein which effects the
?      TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
?      TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
?      NUMBER OF SEQUENCES: 6
?      CORRESPONDENCE ADDRESSES:
?      ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
?      STREET: 1300 I Street, N.W., Suite 700
?      CITY: Washington
?      STATE: D.C.
?      COUNTRY: USA
?      ZIP: 20005-3315
?      COMPUTER READABLE FORM:
?      MEDIUM TYPE: Floppy disk
?      COMPUTER: IBM PC compatible
?      OPERATING SYSTEM: PC-DOS/MS-DOS
?      SOFTWARE: Patent In Release #1.0, Version #1.30
?      CURRENT APPLICATION DATA:
?      APPLICATION NUMBER: US/09/040,444
?      FILING DATE: March 18, 1998
?      CLASSIFICATION:
?      ATTORNEY/AGENT INFORMATION:
?      NAME: O'Connor, Steven P
?      REGISTRATION NUMBER: 41,225
?      REFERENCE/DOCKET NUMBER: 2481.1453-01
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (202)408-4000
?      TELEFAX: (202)408-4400
?      INFORMATION FOR SEQ ID NO: 4:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 1882 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      US-09-040-444-4

Query Match      8.2%; Score 175; DB 3; Length 1882;
Best Local Similarity 47.5%; Pred. No. 1,3e-36;
Matches 585; Conservative 0; Mismatches 640; Indels 6; Gaps 2;

QY      522  ttcacgctgtgacagatgaaatcgtgtgtgtgagacacagtggaagttgccccacc 581
Db      434  TCCCTCATCTCTACTAGTTAACTGCTGTGTGAGAGCGCTGGAAATGACCTTTT 493
QY      582  acctccctgtctcgtgagcggtgtctctcgtcctctcgtgtgtcgggagctgtcaagac 641
Db      494  CAGTCCGTGTGAACCTTGGGCTTCTCTGCGCTCCCTGGTGTGTGATTAATTCAGAGAC 553
QY      642  aggtttggcagagaaagctctctctcgcgaacatgctcgttaacagctgtctcagcttc 701
Db      554  AGGTTTGGCCGTAAAGCTGTCTCTTGTGTGACCAAGCTGGTGCATCTTGTGCGGTG 613
QY      702  ctgcagatttctccatcagcagctggagatgttcaactgtgtatattgtcatctgtggcag 761
Db      614  CTAAACAGCGGTGGCCCAAGCTATACATCATCATGCTCTCTTGGCTGTGACAGGGCAG 673
QY      762  ggccagatctcacaatgtgtgtagcctcactactaactagaacagaaatcttggcaagta 821
Db      674  GTCAGAAAGGCGAGCTGGGTGTCGCGCTATACCTTGATCACAAGAGTTTGTGCGC---TCT 730
QY      822  gtctgattatattctctcaattagagagtgtaacatttttgcaatgtgtctatagctg 881
Db      731  GGCCTACAGAGAAACAGCGGCAATTTGTATACAGATGGCCCTTCACAGTGGGCTACTGGGG 790
QY      882  ctgcacatgttcttactactacagacatgcagatgcgtctgctgcgcgtgacgtgacgtg 941
Db      791  CTTGGCGGGGTGGCTATGCCATTCACAGACTGGCCCTGGCTCCAGCTAGCTGTGTCCTG 850
QY      942  ccggagagtgctgtgtctcccgctgtgtgtgttcaatctcctgaatctcccgatgctgata 1001

```









QY 1286 tgaattccagcttaacatcacagcctggctgctatgtcgaaagctgcccagcgtat 1345  
1214 TGACATTCGCCGCAAGTTCATGACAAATCCTCCATTAAGTATGCTGGGCGGCCGCAATCAC 1273  
QY 1346 catagctcgaagtatgtcttggggaggggtgtgtcttcttcaatcaactgttacctgt 1405  
1274 TCAGGCGCTTCCTGATCCTGCGAGAGTGGCCATCCTGCGCCTCATCTTTGTGCTTC 1333  
QY 1406 ggaattactcttaccatgtctgtcatgtcgaggaaatttggaaacttggaaactctgc 1465  
1334 AGAAATGACGCTCTTGAGAAACAGCACTGGCTGATTTGGGAAAGGAAAGCCGTGTGGCTC 1393  
QY 1466 ttctccatgctgtatgtcttcaactgcgtgagcttaccacccctgttcaaggaaatgcgc 1525  
1394 CTTCACCTCCCTCTTCTCTTACCAAGAGAGCTTACCCCTACAGTCCCTCAGGCAAAAGG 1453  
QY 1526 ggtgggggtcacatcacagcgcctccagagttggcagacatcatgcccctacttgttta 1585  
1454 TATGGTATCATGATACATATGSGCTCGAGTGGGAAGTATGATACCCCACTGGTGAAAT 1513  
QY 1586 cctgggtcttacaacagaatgcgcctcacaatgctgtagtgtagtgcgtctgtat 1645  
1514 CACGGGAGAACTGACAGCCCTTCATCTAATGTCACTTTTGGACCATGACTCTACCTG 1573  
QY 1646 tgaatcttcacccttttctccctgaaagtggaaattggaactcttccagaacttga 1705  
1574 AGGCACTGCTGCTCTTCTCTTGTGAGACCCCTCAATGCGCCCTTACCAAGAAATATGCA 1633  
QY 1706 gcaagtga 1714  
Db 1634 GGACATACA 1642

RESULT 8  
US-08-592-126-108/C  
; Sequence 108, Application US/08592126  
; Patent No. 5821091  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
; TITLE OF INVENTION: Polypeptides  
; NUMBER OF SEQUENCES: 151  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,126  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 108:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 370 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO

; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: 678.seq  
US-08-592-126-108

Query Match 4.9%; Score 104; DB 1; Length 370;  
Best Local Similarity 76.2%; Pred. No. 3.2e-18;  
Matches 128; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 641 caggtttgagggaggaagacttcttcttcgacacatgctgtaagcttgcttgaact 700  
Db 221 CAGGTTTGCCCGAAGAAATGCTGTTGTGACCATGGCGCTGCGAGAGCTTCACTT 162  
QY 701 cctgaagatttctccatcacagcttggagatgttcaactgttatttgcacgttggacat 760  
Db 161 CCTGACAGACCTTCTGAGAAATTTGAGATGTTGTGCTGCTGTTTCTCTGTAGGCAAT 102  
QY 761 gggcagatcttcaactatgtgttagccttacttacttggaaagaaat 808  
Db 101 GGGCGAGATCTTCACATATGGGGCAGCATTTGTCTGGGATGCGCAAT 54

RESULT 9  
US-08-592-126-109  
; Sequence 109, Application US/08592126  
; Patent No. 5821091  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Dolganov  
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
; TITLE OF INVENTION: Polypeptides  
; NUMBER OF SEQUENCES: 151  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,126  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 481 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: H993.seq  
US-08-592-126-109

Query Match 4.9%; Score 104; DB 1; Length 481;  
Best Local Similarity 76.2%; Pred. No. 3.8e-18;  
Matches 128; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

01 641 cagatttggaggaagaacgctctctcgcacaacatgctgtacagactgcttcaagctt 700
02 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
03 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
04 Db 169 CAGGTTTGCGGGAGAGAAATGCTGTTCTGGCCATGGGCATGCAAGACGCTTCAGCTT 228
05
06 Qy 701 cctcgcagatttctccatcagctgaggaatgttcactgtgtatttgcactgtggagat 760
07 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
08 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
09 Db 229 CCGTGAAGATCTTCTCGAAGAAATTTTGAGATGTTTGGCGTGCTGTTTGGCTTTGATGAGCAT 288
10
11 Qy 761 gggccagatctccaactatgtgttagccttcatactaggaacagaat 808
12 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
13 ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
14 Db 289 GGGCGAAGATCTCCAACTATGTGGACAGCATTTGTCCGGGTATGGGCAT 336
15
16 RESULT 10
17 PCT-US95-14418-3
18 Sequence 3, Application PC/TUS9514418
19 GENERAL INFORMATION:
20 APPLICANT:
21 APPLICANT:
22 TITLE OF INVENTION: DNA Encoding a Thermostable DNA Polymerase Enzyme
23 NUMBER OF SEQUENCES: 51
24 CORRESPONDENCE ADDRESS:
25 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
26 STREET: 6300 Sears Tower, 233 South Wacker Drive
27 CITY: Chicago
28 STATE: Illinois
29 COUNTRY: United States of America
30 COMPUTER READABLE FORM:
31 MEDIUM TYPE: Floppy disk
32 COMPUTER: IBM PC compatible
33 OPERATING SYSTEM: PC-DOS/MS-DOS
34 SOFTWARE: PatentIn Release #1.0, Version #1.25
35 CURRENT APPLICATION DATA:
36 APPLICATION NUMBER: PCT/US95/14418
37 FILING DATE:
38 CLASSIFICATION:
39 ATTORNEY/AGENT INFORMATION:
40 NAME: Gass, David A.
41 REGISTRATION NUMBER: 38,153
42 REFERENCE/DOCKET NUMBER: 28003/32330
43 TELECOMMUNICATION INFORMATION:
44 TELEPHONE: 312/474-6300
45 TELEFAX: 312/474-0448
46 TELEX: 25-3856
47 INFORMATION FOR SEQ ID NO: 3:
48 SEQUENCE CHARACTERISTICS:
49 LENGTH: 1794 base pairs
50 TYPE: nucleic acid
51 STRANDEDNESS: single
52 TOPOLOGY: linear
53 MOLECULE TYPE: DNA (genomic)
54 FEATURE:
55 NAME/KEY: CDS
56 LOCATION: 1..1794
57
58 PCT-US95-14418-3

```

Query Match	2.6%;	Score 55;	DB 5;	Length 1794;
Best Local Similarity	46.1%;	Pred. No. 7.5e-05;		
Matches 184;	Conservative	0;	Mismatches 215;	Indels 0;
			Gaps	0

Accession	Sequence	Length
QY	187 gggggcccttcagcgcgtcaatctctctctgctcagcgccagcatatctcccaatggtc	246
Db	404 gggagcagcccatctctctctccctaccctccttgagaccctctcaacaccaccccgagagggg	463
OY	247 tcaatgtaatctcaagtctgttctctgcgggagcccgagacaacgctctgtcagtgtagcgg	306
Db	464 tggcccgccgcttaccgggggggagtgtagcagggagcgcgcgcacccgggcccctctcgcg	523
OY	307 acgcgcggaactctgagcagcgccctgcygcacaacacagtgtccgcctctcggctcggagcg	366
Db	524 agagccttcattcggaaacctctttagcgccctcaggggggagagaaacacctctcttggctct	563

QY	367	gcccgaagctgccccacagctcgagccgctacggctcgccacatccgcaacttcg	426
Db	564	ACCCAGAGGTGGAAAAGCCCTCTTCCCGGGCTCTGGCCCAATGGAGGCCACCGGGTAC	643
QY	427	cgctcgagctlgagccggcgcgcgctlgagcctlgagcctlgagcagagagctgcgc	486
Db	644	GGCTGGAGCTGGCCCTACCTGGAGGCCCTTTCCTCGGAGCTTGGGGAGAGATCCGCGCC	703
QY	487	tggatgctctgtagtctcaagccagagcgtctactcgttccaccgctcgagaccgagtgtgattc	546
Db	704	TCCGAGGAGAGGACTTCCGCTTGGCGGGCCACCCCTTCAACTCACTACCTCCGGGACCAAGC	763
QY	547	tggctgtgtagagacaactcgagagctgcgccctccaccact	585
Db	764	TGGAAAGGCTCTTTGACAGAGCTTAGGCTTCCCGCCT	802

```

11 RESULT 11
12 PCT-US95-15327-3
13 : Sequence 3, Application PC/TUS9515327
14 : GENERAL INFORMATION:
15 : APPLICANT:
16 : TITLE OF INVENTION: Biologically Active Fragments of
17 : TITLE OF INVENTION: Thermus Flavus DNA Polymerase
18 : NUMBER OF SEQUENCES: 51
19 : CORRESPONDENCE ADDRESS:
20 : ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
21 : STREET: 6300 Sears Tower, 233 South Wacker Drive
22 : CITY: Chicago
23 : STATE: Illinois
24 : COUNTRY: United States of America
25 : COMPUTER READABLE FORM:
26 : MEDIUM TYPE: Floppy disk
27 : COMPUTER: IBM PC compatible
28 : OPERATING SYSTEM: PC-DOS/MS-DOS
29 : SOFTWARE: Patentin Release #1.0, Version #1.25
30 : CURRENT APPLICATION DATA:
31 : APPLICATION NUMBER: PCT/US95/15327
32 : FILING DATE:
33 : CLASSIFICATION:
34 : ATTORNEY/AGENT INFORMATION:
35 : NAME: Gass, David A.
36 : REGISTRATION NUMBER: 38,153
37 : REFERENCE/DOCKET NUMBER: 28003/31716
38 : TELECOMMUNICATION INFORMATION:
39 : TELEPHONE: 312/474-6300
40 : TELEFAX: 312/474-0448
41 : TELEX: 25-3856
42 : INFORMATION FOR SEQ ID NO: 3:
43 : SEQUENCE CHARACTERISTICS:
44 : LENGTH: 1794 base pairs
45 : TYPE: nucleic acid
46 : STRANDEDNESS: single
47 : TOPOLOGY: linear
48 : MOLECULE TYPE: DNA (genomic)
49 : FEATURE:
50 : NAME/KEY: CDS
51 : LOCATION: 1..1794
52 PCT-US95-15327-3

```

Query Match	2.6%;	Score 55;	DB 5;	Length 1794;
Best Local Similarity	46.1%;	Pred. No. 7.5e-05;		
Matches 184;	Conservative	0;	Mismatches 215;	Indels 0;
			Gaps	0;

OY	187	gggggccccttcagagcctaatctctcttcgctcagagccagatcatcccaatggct	246
DB	404	gggagagaccctactctctctccgcttacccttccgagaccctccaaacacacccccgaagggg	463
OY	247	tcaatgtatgttcagtcgtgttctctcgagcgagaccgccgagacacgcgtcttcgagttccgg	306
DB	464	tggccggcgcttaccggggggagtgagagagagacggccggccacacggggccctctcttcgg	523

Query Match	2.6%;	Score 55;	DB 5;	length 3048;
Best Local Similarity	46.1%;	Pred. No. 0.0001;		
Matches 184; Conservative	0;	Mismatches 215; Indels	0;	Gaps 0;

```

QY 187 gggggccctcccaagcgcctcatcttctctcgtcagcgccagcatcatcccaatgct 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1415 GGGAGCAACCCATGCTCTGCGCTACCTCTGAGACCCCTCAACACCCACCCGAGGGG 1474
QY 247 tcaatgtatgtcagtgtctgtctctgagggagcccgagaccggtgtcagatgccc 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1475 TGGGCGCGGCTACGGGGGGAGTGGAGAGAGAGCCGCCACCGGCCCCCTCTCTCGG 1534
QY 307 acgcgcgaacctgagcagcgccctgcgcgaacaacagtgtcccgctgcggtgcggagc 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1535 AGAGGCTCCATCGGACCTCTTAAGCGCCCTCGAGGGGGAGAGAACCTCTTGGCTCT 1594
QY 367 gcccgcgaagtgtcccaacagctgcagcgtacgcggtctgcgcacatgcgcaactctccg 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1595 ACCAGAGGTGGAAGAGCCCTCTCCCGGCTCTGGGCCCATGAGAGGCCACCGGGGTAC 1654
QY 427 cgtcggcgtgagcgcggcgagcgtgcagcgtgcggcgcaagtgcagcagagcgtgc 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1655 GGCTGACGTGCTACTGCTGAGGCCCTTTCTCTGAGCTGCGAGAGAGATCCGGCGCC 1714
QY 487 tggatgctgggagctcagcagcagcgtctacgtgtccaccgctgcagcagtgagtc 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1715 TCGAGAGAGAGGTCTTCGGCTTGGGGGCCACCCCTTCACCTCAACTCCCGGAGCACAGC 1774
QY 547 tgggtgtgagagacaactcgtgaagtgccccctcacacct 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1775 TGGAAAGGTGCTCTTTGACGAGCTTAGGCTTCCCGCT 1813

```

```

RESULT 14
US-08-648-657-14
; Sequence 14, Application US/08648657
; Patent No. 5885813

```

```

; GENERAL INFORMATION:
; APPLICANT: Davis, Maria
; APPLICANT: Moffett, R. Bruce
; APPLICANT: Fuller, Carl W.
; TITLE OF INVENTION: THERMOSTABLE DNA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/648,657
; FILING DATE:

```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: one
; APPLICATION NUMBER: US 08/455,686
; FILING DATE: May 31, 1995

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Walburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 219/304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1686 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: FY4
; LOCATION: 1...1683
US-08-648-657-14

```

```

Query Match 2.5%; Score 53.4; DB 2; Length 1686;
Best Local Similarity 45.9%; Pred. No. 0.00019;
Matches 183; Conservativity 0; Mismatches 216; Indels 0; Gaps 0;

```

```

QY 187 gggggccctcccaagcgcctcatcttctctcgtcagcgccagcatcatcccaatgct 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 GGGAGCAACCCATGCTCTGCGCTACCTCTGAGACCCCTCAACACCCACCCGAGGGG 355
QY 247 tcaatgtatgtcagtgtctgtctctgagggagcccgagaccggtgtcagatgccc 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 TGGGCGCGGCTACGGGGGGAGTGGAGAGAGAGCCGCCACCGGCTCTCTCGG 415
QY 307 acgcgcgaacctgagcagcgccctgcgcgaacaacagtgtcccgctgcggtgcggagc 366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 AGAGGCTCCATCGGACCTCTTAAGCGCTCGAGGGGGAGAGAGCTCTTTGGCTCT 475
QY 367 gcccgcgaagtgtcccaacagctgcagcgtgcgcgtgcgcacatgcgcaactctccg 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 ACCAGAGGTGGAAGAGCCCTCTCCGGGCTCTGGGCCCATGAGAGGCCACCGGGGTAC 535
QY 427 cgtcggcgtgagcgcggcgagcgtgcagcgtgcagcgtgcagcagtgagagagctgct 486
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 GGCTGAGAGGTGCTACTGCTTACAGGCCCTTTCTCTGAGCTGCGAGAGATCCGGCGCC 595
QY 487 tggatgctgggagctcagcagcagcgtctacgtgtccaccgctgcagcagtgagtc 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 TCGAGAGAGAGGTCTTCGGCTTGGGGGCCACCCCTTCAACCTCAACTCCCGGAGCACAGC 655
QY 547 tgggtgtgagagacaactcgtgaagtgccccctcacacct 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 656 TGGAAAGGTGCTCTTTGACGAGCTTAGGCTTCCCGCT 694

```

```

RESULT 15
US-08-073-384C-3
; Sequence 3, Application US/08073384C
; Patent No. 5541311

```

```

; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamachev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVENSTOCK, MEDLEN & CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/073,384C
; FILING DATE: 04-JUN-1993

```

```

; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2002, 19:33:44 ; Search time 4179.87 Seconds  
(without alignments)  
6893.994 Million cell updates/sec

Title: US-09-521-195b-2

Perfect score: 2135  
1 ccccgcttcgcgcgcccaat.....aactactccaataaataat 2135

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
C 1	918.4	43.0	928 9	AL557808 AL557808
2	883.4	41.4	905 9	AL557809 AL557809
3	362.6	26.4	800 10	B1683731 603306292
4	535.6	25.1	910 10	B1552541 603194456
C 5	528.8	24.8	559 9	AI032763 ox13h02.x
6	521.2	24.4	613 9	AM264315 xq98b07.x
C 7	514	24.1	577 9	AM476024 uq93d03.x
8	505.4	23.7	544 9	AM592711 bfa6e06.x
C 9	496.4	23.3	498 9	AI628726 ty78b05.x
10	490.6	23.0	611 10	B6723644 602697810
11	488.8	22.9	879 9	B6870249 602791116
12	466.8	21.9	470 9	AA428395 zva48b07.r
13	457.2	21.4	754 10	B6332869 602430714
14	450.6	20.5	799 10	B1647860 603276129
15	438.4	20.4	578 10	BE846144 232219 BA
C 16	435	20.1	471 9	AM450142 UI-H-B13-
17	430	20.1	460 9	AI631946 wa38a05.x

C 18	429.4	20.1	507 9	AM342047
19	428	20.0	806 10	B1461374
20	423.4	19.8	846 10	BF783848
C 21	422	19.8	433 10	N26836
22	422	19.8	917 10	BF785653
23	419.8	19.7	641 9	BB615168
24	416	19.5	640 9	BB615170
C 25	414	19.4	442 9	AI130318
26	414	19.4	689 10	B1547770
27	413.6	19.4	514 10	BE846351
28	411.8	19.3	839 10	BF784044
C 29	407.8	19.1	426 9	AM296942
30	394.8	18.5	645 9	BB617187
31	392.4	18.4	522 10	BF230186
C 32	392	18.4	605 9	BB614876
C 33	387.2	18.1	556 9	AI785332
C 34	384	18.0	402 9	AA425843
C 35	381	17.8	403 10	H93112
C 36	379.8	17.8	922 10	B1251565
37	379.6	17.8	866 10	B6361802
38	374	17.5	482 10	BM032317
39	373.8	17.5	484 10	BF774599
40	373.2	17.5	762 10	BF782868
41	367.6	17.2	456 10	H93240
C 42	354	16.6	377 9	AA699791
C 43	352	16.5	371 9	AA953556
44	348.4	16.4	606 10	BF152158
45	348.8	16.3	371 10	B1764248

## ALIGNMENTS

RESULT 1  
AL557808/c 928 bp mRNA linear EST 16-FEB-2001  
LOCUS AL557808 LTI\_NFL008\_TC2 Homo sapiens cDNA clone CS0DDU003YB03 3  
DEFINITION prime, mRNA sequence.  
ACCESSION AL557808  
VERSION AL557808.1 GI:12901773  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 928)  
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

## FEATURES

source

1..928  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DDU003YB03"  
/clone\_id="LTI\_NFL008\_TC2"  
/sex="male"  
/tissue\_type="T cells from T cell leukemia"  
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax: (1) 301 610 8371  
Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com"  
BASE COUNT 264 a 196 c 208 g 252 t 8 others  
ORIGIN





```
|||||
Db 441 TCTGCGGCTCGGCTGAGACCGGCGCGACGTGACCTGGGGCAGCTGGAGCAGAGAGA 500
Qy 481 gctgcctgagatgctggaattcaagccagagcttaacctgtcacacgctgtagccgagt 540
Db 501 GCTGCTGGAGATGGCTGGAGATTCAGCCAGAGCTTACCTGTCCACCGTGTGACCGAGT 560
Qy 541 ggaatcgtgtgtgagagacaactggaaggtgccctcacacacctccctgtcttcgtag 600
Db 561 GGAATCTGGTGTGAGAGACAACAGAGAGTGCCTCCACACCTCCCTTCTTCTGCTAG 620
Qy 601 gcatgctcctcgctcctcgtgtgcggagagctgtcaagcaaggttggcagaagaagc 660
Db 621 GCGTCTCTCGGCTCCTTCTGTCTCCGGGACGCTGTCAAGAGTGTGACAGAGAAGACG 680
Qy 661 ttctcttcgcaacatgctgtagaagactgtctcaagctccctcgcaagattttccatca 720
Db 681 TTTCTCTTGGCAACCATGGCTGTACAGACTGGCTTCACTTCTGCAATTTTCTCCATCA 740
Qy 721 gctggagagatgtcaactgtgtatttcatctgtgagcatgagccagatcccaactatg 780
Db 741 GCTGGAGAGATTTCACTGTGTATTGTTCATTCGTGGGCAATGGGCAATCTCCAACTATG 800
Qy 781 tgttagccttcaactagtagaacaagaattctgtgcaagtaagttcgtattatctcta 840
Db 801 TGGTAGCCTTCACTAGTAAGAAAGAAATCTTGCAAGTCAATCTGATTAATATCTCTA 860
Qy 841 cattagagatgtgacatttttgcagttgcatatgctgcgtgc 885
Db 861 CATTAGAGATGTGCACATTTTTCAGATGTGCTATATGTTGCTTC 905
```

```
RESULT 3
BI683731 800 bp mRNA linear EST 18-SEP-2001
LOCUS BI683731
DEFINITION 60306292F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:5342279 5',
mRNA sequence.
ACCESSION BI683731
VERSION BI683731.1 GI:15646359
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 800)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apds@email.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LRAM1869 row: h column: 24
High quality sequence stop: 793.
Location/Qualifiers
1. 800
```

```
FEATURES
Source
1. 800
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5342279"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
```

```
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
```

```
BASE COUNT 150 a 226 c 227 g 197 t
ORIGIN
```

```
Query Match 26.4%; Score 562.6; DB 10; Length 800;
Best Local Similarity 81.2%; Pred. No. 2e-111;
Matches 693; Conservative 0; Mismatches 99; Indels 61; Gaps 1;
```

```
Qy 143 aagcatgaggagctcagagaggtatgccttccttgaggaggtggagcccttcacag 202
Db 1 AAACATGCGGAGCTACGACAGAGTATGCTTCTG6GCGAGTGG6GCTTTCACAG 60
Qy 203 cctcatctcttccttgctcaagcagcaatccccaatgcttcaatgtagttagt 262
Db 61 CCGCATCTTCTTCTGCTCAGGCGCAGCATATCCCAATGGCTTCATATGATATTCAGT 120
Qy 263 cgtgtcctgaggagagcccgagacccgtgtcagatgcccgaagcccgaccttag 322
Db 121 CGTGTCTCTGGGCGGAGCCCGAGACCGTGTGCTGCTGCTGACATGTGAACCTGAG 180
Qy 323 cagcgcttgagcaacaacagatgtcccgctgagctgagcagccgaggtgcccac 382
Db 181 CAGCTCTTGCGAACACACAGCATCCCTTGGAGAGAAAGACGACACAGCTCTCA 240
Qy 383 cagctcagccgctacagcgtccgcacacatccgaactctcagcgtcggagctgagac 442
Db 241 GACCTCCGCGCGTACCGATGCGACATGCGCAACATCTCTGGGATGGGCTGAGCC 300
Qy 443 ggggagcagctgagacctgaggagcagctgagagagagagctgagatgctgagatt 502
Db 301 AGGACAGAGACGTGATCTTGAGACAGCTGAGAGAGAGACCTCTGATGAGCTGGAGTA 360
Qy 503 cagccagagcgtctaccctgtccaccgtcgtagccagatgtaatcgtgtgtagacaa 562
Db 361 CGACAAAGACATCTTCTCTGCAACATCGTAGAC----- 395
Qy 563 ctggaagtgcccctcaacacactccctgtcttcgtagagcgtgctcctcgtcgt 622
Db 396 -----AGCGTTCTCTCGGCTCTTGT 419
Qy 623 gtccggagcagctgtcagaagatttgcaaggaagaagttccttcctcgaacatgctgt 682
Db 420 GTCTGGGCAAGCTCTGACAGAGTTTGGCAGGAAGAAAGTCTTTCACCAATGGCTGT 479
Qy 683 acagactggttcaagcttccctgagatttcccatcagctgaggagatgttaactgtgt 742
Db 480 GCAAGACTGGAATTCAGCTTCTGTGAGATTTTCTCAACCACTGGAGATTTCACTGTGT 539
Qy 743 attgtcatcgttgagcatgagccagatcccaactatgtgttagccttcaactaggaac 802
Db 540 GTTTGGCATGTGGGAGATGGGCGAGATTCACAACTACGTGAGTGGGCTTCACTAGAAC 599
Qy 803 agaaatctctggcaagtcagatcgtattatattctcaatagagaggtgtgcaatttt 862
Db 600 TGAATCTCTAGAGAAATCGGTGGATTCATCTTCCATTAAGAGTGTGATCAATTTT 659
Qy 863 tgaagtgtgctatagtctgtgcacatgtttgttacttaacatcaagagactgagatgt 922
Db 660 TGCATATGGCTACATGCTCTGCGGCTGTTTGATATCTTCATCAGAGACTGAGATGCT 719
Qy 923 gctgtcgtcgtcgtcgtgagcgtgagtgctgtgtgtccgctgtgtgttcaattcctga 982
Db 720 GCTGCTGGCCCTGACACTGCTGCGCTGTGTGTTCCTGCTGTGTGTATTCACAGA 779
Qy 983 atctcccgatgtg 995
Db 780 ATCTCCCGGCTGG 792
```

```
RESULT 4
BI552541 910 bp mRNA linear EST 05-SEP-2001
LOCUS BI552541
```

DEFINITION	603194456t1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5274108 5', mRNA sequence.
ACCESSION	B1552541
VERSION	B1552541.1 GI:15439853
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homimidae; Homo.
AUTHORS	1 (bases 1 to 910)
TITLE	NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki Toshiyuki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov Plate: LHAM1691 row: p column: 13 High quality sequence stop: 671. Location/Qualifiers 1..910
FEATURES	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5274108" /clone_lib="NIH_MGC_95" /tissue_type="hippocampus" /lab_host="DH10B" /note="Organ: brain; Vector: pBluescriptR (modified Bluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgagaa...) Oligo-dn primed using primer 5'-TTTCTTTTCTTTTTCVN-3', size-selected for average insert size 2.5 kb and normalized to R0F 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT	131 a 284 c 289 g 206 t
ORIGIN	
Query Match	25.1%; Score 535.6; DB 10; Length 910;
Best Local Similarity	88.1%; Pred. No. 1.5e-105;
Matches 583:	Conservative 0; Mismatches 79; Indels 0; Gaps 0;
Db	
QY	129 ggaacgcgacgtgggaagaatgcggaacacgaaggatcgccttccttgcgagctg 188
Db	
QY	5 GGGGCTGTGAGGGGGCGCATGCGGGACTACGACGAGTGATGCCCTTCTGGCAATGG 64
Db	
QY	189 ggagcccttcagcgactattttctctgctcagcgaccagcatatcccgaatgcttc 248
Db	
Db	65 GGGCCCTTCCAGCGCCTATCTTCTCTGCTCAGCGCCACATCATCCCAATGGCTTC 124
QY	249aatgatgtaagtctgtttctctgscgggccccggagcaacgctgtcagatgcggac 308
Db	
Db	125 ACGGCGCTGCTCCGCTTCTGTATACGACCCCAGGACACCGCTGCGGGATCGGGAC 184
QY	309 gcgcgacaactgaagcagcgctctggcgaacaacaaagttcccgctctgcgtagcagc 368
Db	
Db	185 GCCCGAACCTGAGACAGCGCTGGGGCAACAACACTGTCCACTGCGGCTCGGGACGGC 244
QY	369 cgcgaagtgcccaacagctcagccogtaaccgctcgcacacatcgccaactctcgagc 428
Db	
Db	245 CGCGAGGTGCCCAAGCTGCGCGGCTACCGGGCTCCGACCACATGCGCAACTTCGCGCG 304
QY	429 ctctggagctgagccggggcgagactgtgaccttgggcgagctcgagacagagaagctgctg 488
Db	
Db	305 CTCGGGCTGGAGCCGGGGCGCAGCTGACCTGGGGCACGCTGGAGCAGAGAGCTGTCTG 364

OY	489	gatgctgtggagatcctcaaggacgcgtctactcgtccacgcgttgtcacgcagtgtgacgcagttgaaactctg	548
Db	365	GATGGGTGGGAGTTCAGTAGTGAAGACGCCTCACTCGTCCACCATTGTGTGACCGAAGTGGAACCTG	424
OY	549	gtctgtgaagaacaacgcgggaaggtgcccccaccacaccttcctgttctcttgtaagcgctgctc	608
Db	425	GTTGTGTGAGGACGACTGGAAGAGGCCCACTCACAAATCTCCTTGTCTTCTGTTGGGTGTGCTG	484
OY	609	ctcgagctcctctgltgcgcggcagcgtctgacaagaagtttgcagagaagaacgttctcttc	668
Db	485	TTCGGCTCCTCATTTTCAGGGCAGCTGTGCAGACAGSTTTTGC CGGAAGAATGTGCTTC	544
OY	669	gcaaccatgtgcgtgtaaacagctgtctcaactcttcctgcagaatttctcctcagcgtggag	728
Db	545	GTCACATCTGGCATGGAAGACAGCGCTTCACCTTCCTGCACAATCTTTCGGAAGAAATTTTGGAG	604
OY	729	atgttacacgtgttatcttcatctcgtggcattggcgcagatctccaacatgttgtgtgacc	788
Db	605	AATGTGTGTGCTGCTGCTTTTCTCTGTGAGCATTTGGCCAAATCTCCAAGTAATGTGGACGCC	664
OY	789	tt 790	
Db	665	AT 666	
RESULT	5		
LOCUS	AI032763/c		
DEFINITION	AI032763	559 bp mRNA linear EST 28-AUG-1998	
ORIGIN	ox13h02.x1 Soares_fetal_liver_spleen_INFLS.S1 Homo sapiens cDNA clone IMAGE:1656243 3' similar to TR:O15456 O14546 POLYSPECIFIC ORGANIC CATION TRANSPORTER., , mRNA sequence:		
ACCESSION	AI032763		
VERSION	AI032763.1	GI:3253460	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 559)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Stransberg, Ph.D. Email: csapsb-remail.nih.gov This clone is available royalty-free through DNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1014 Std Error: 0.00 Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop: 353. Location/Qualifiers		
FEATURES	Source		
	1..559		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_xref="IMAGE:1656243"		
	/clone_lib="Soares_fetal_liver_spleen_INFLS.S1"		
	/sex="male"		
	/dev_stage="20 week-post conception fetus"		
	/lab_host="DH10B (ampicillin resistant)"		
	/note="Organ: Liver and Spleen; Vector: pTV73D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5'] AACGGAAGAATTAATAAGATCTTTTTTTTTTTTTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT	151 a	95 c	110 g 203 t
ORIGIN			

Query Match	24.8%;	Score 528.8;	DB 9;	Length 559;
Best Local Similarity	98.7%;	Pred. No. 3.9e-104;		
Matches 533; Conservative	0;	Mismatches 7;	Indels 0;	Gaps 0;

OY	1596	taaacgaatgctgcctcacatcgcgcacgggtacgtcagctgcctattggaatctc	1655
Db	559	TTTCAGCAGAAATGCTGCCATCATCATCGTCATGGGTAGTCTACTGTCTCTATTGGAAATCCTC	500
OY	1656	accctttttccctgaaagtlttggaatgactctccagaaccttagagcaatgacag	1715
Db	499	ACCCCTTTTTTCCCTGAAAGTTTGGGAATGACCTCTCCGAAAGACCTTTAGACGAGATCGAG	440
OY	1716	aaatggaatggtttgagatctgggaaaaaaacaagagctcaatggaaacgaanaaat	1775
Db	439	AAATGGAATGgtTCACATCTGGGAAAAAACAGAGACTCATGTGAGACACAAAGAAAT	380
OY	1776	cccaaggttcttaatbaactgcatcttgaaaaaaatatcacccatttgytgaagtgaanaa	1835
Db	379	CCCAAGGTTCCTTAATACGTGATCTGAAAAAATATCTCCCATTTGGTAAGTGAANA	320
OY	1836	cagaaaaataagaccctgtgagaaatctgtgttccacctaataatgagctgacgtac	1895
Db	319	CAGAAAAATAGACCCGTGAAGAATTTGTTGTTCCACTGAATAGGAGCTAGCTGAAC	260
OY	1896	gattgacacccaatgagacctgtctacaagaatgctgctacatacagtaaaccttgat	1955
Db	259	GATTGACACCAAAATGAACTTGCTGTATCAAGAAATGCTGTATACAGTAATCTGTGAT	200
OY	1956	gattcttcagataatgtcctgtccttacaacaacacattcttagagagtctcctact	2015
Db	199	GATTCTTCCAGATATATCTCTTATATTACAAACCAACATTTCTGAGAGTCTCTTACT	140
OY	2016	cattaatccaatgaaatggaatgtgttaagatgctcttgaacaatgtaagcaagactggt	2075
Db	139	CATTAAATTCATGAATATGATGTGGTAAGATGCTTGAAACAATGTTAGTCAMGAGCATG	80
OY	2076	aaaatacatataaagattacaactcatlttccaatcatacaaatcatatccaataaaaat	2135
Db	79	AAAATCATATATAAGATTAACCTCATTTCTCCATCTATACAAATATCAATCAATATAAAT	20

RESULT	6
AM264315/c	
LOCUS	
DEFINITION	AM264315 613 bp mRNA linear EST 28-DEC-1999
	xg98b08.x1 NCI CGAP Brn53 Homo sapiens cDNA clone IMAGE:7758647 3'
	similar to TR:014546 O14546 POLYSPECIFIC ORANAGNIC CATION
TRANSPORTER :	/ mRNA sequence.

JOURNAL  
Unpublished (1998)  
COMMENT  
Contact: Robert Strausberg, Ph.D.

Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: [www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco

High quality sequence stop:	415
Location/Qualifiers	
1. .613	

BASE COUNT	163	a	112	c	133	g	205	t
ORIGIN								

Query Match	24.4%	Score 521.2	DB 9	Length 613
Best Local Similarity	98.5%	Pred No. 1.8e+102		
Matches 526	Conservative 8	Mismatches 0	Indels 0	Gaps 0

QY	1602	agatagtcgcccatacgcgcacatgggaagctgaagctgctcattggaaatcttaacctt	166
Db	553	AGTAGTCGCCCCCACCATTCGTCATGGGAGTGAAGCTGACTGTCTCTATTGGATCCCTTACCCTT	494
QY	1662	ttttcccttgaagaatttgggaatgaactcttcagaaaccttagagacagatgcagaagt	1721
Db	493	TTTTTCCCTCGAAGCTTTGGGAATGACTCTTCCGAAACCTTTAGAGCAGATCGCGAAATGG	434
QY	1722	aaatggttcacatctgggaaaaaaacaagaagactgaatggagacagaagaataatccgaag	1781
Db	433	AAATGTTTCACATCTGGGAAAAAACCAAGAGCTTCATGGAGACGACAGAAATCCCAAG	374
QY	1782	gtctcaataacatgaactctgaaaaaataatactaccccatcttggatgaaatgaacaagaa	1841
Db	373	GTTCTAATAATCGATTCTCGAATAAATATACCCATTGGTGAAGTGAATAAACAGAAA	314
QY	1842	aataagaccctgggaaaaaatttggtttcccaactaaatlgagctgaagctgaacattga	1901
Db	313	AATAAGACCCTGGTGAGAAATCTGTTTCCACTGAATAAGGACTGACTGTAACCATTTGA	254
QY	1902	cacccaaatgaaccttgcatacagaagaatgctgctatatacagtaaacctctgtagatctt	1961
Db	253	CACCAAAATGGAACCTTGCTATCAAGAANAATCTGTCATACAGTAACCTGTGATGTTCT	194
QY	1962	tccagataatgtcttgcgtttacaacaacacacacttcttagagagttcctctactaataa	2021
Db	193	TCCAAATATATGCTCCGTGCTTTACAAACCAACCACTTCTTGAAGGTGCTCTTACTCATPAA	134
QY	2022	ttcaatggaatggaatgttggaagatgctcttgaagaacatgctgaagcaagactggtlaaata	2081
Db	133	TTTCATGTGAATGGATGTGGTAAAGATGCTTTGAAAACATGTTAGTCACAGGACTGTGTAATAA	74
QY	2082	catataagaattaacacactatcttccaatctaaacataaataactatccaaataaaat	2135
Db	73	CATTATAAAGATTAAACGCTCATTTTCCAACTATACCAAAATACTCATTCCAATATTAATAAT	20

LOCUS	DEFINITION	TRANSPORTER
AA476024/c	577 bp mRNA	linear
AA476024	ug84d03.x1 NCI CGAP lu33 Mms musculus cDNA clone IMAGE:5937701 3'	EST 24-FEB-2000
	similar to TR:O14546 O14546 POLYSPECIFIC ORGANIC CATION	
	TRANSPORTER. ;, mRNA sequence.	

JOURNAL  
COMMENT

Tumor Gene Index  
Unpublished (1997)  
Other ESTs: ug83d03.x1  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
www.bio.lnlnl.gov/dbtrp/image/image.html

MG1:1050121  
Seq primer: -40UP from gibco  
High quality sequence stop: 208.  
Location/Qualifiers

1. 577  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2937701"  
/clone\_lib="NCI-CGAP\_Lu33"  
/tissue\_type="Pooled lung tumors"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site.1: NotI; Site.2: EcoRI; 1st  
strand cDNA was prepared from mRNA obtained from pooled  
lung tumors with a Not I - oligo(dT) primer [5'  
TGTTACCATCTGAAGTGGAGCGGCCGCTCTGTTTCTTTTCTTTT 3']  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 157 a 98 c 114 g 208 t  
ORIGIN

Query Match 24.1%; Score 514; DB 9; Length 577;  
Best Local Similarity 97.1%; Pred. No. 6.3e-101;  
Matches 534; Conservative 0; Mismatches 15; Indels 1; Gaps 1;  
QY 1587 ctccggtgcttaacaagaatgctgctcctacatcgtcatggtgtagtctgactgctgatt 1646  
|||||  
DB 577 CTGGGTGCTTACACAGATGCTGCCCTACATCGTCATGGGAGTCTGCTCTGTGAT 518  
QY 1647 ggaatctcaccctcttttccctgaaagtttggaatgactcttcagaaacctagag 1706  
|||||  
DB 517 GGAATCTCACCCCTTATTTTCCCTGATAGTGGAGAAAGACTCTTCCAGACACCTTAGAG 458  
QY 1707 cgaatgcagaagtgtaaatggttcagatctggg-aaaaaacaagagactcaatggaagac 1765  
|||||  
DB 457 CAGATGCAGACAGTGAATGTTGATCTGGGTAAATAATCAATAGTTCATAGTACAC 398  
QY 1766 agaagaaaaatcccaaggttctaataactgcaatcttgaaaaaataatctacccattggtg 1825  
|||||  
DB 397 AGAAGAACATCTCAAGGTCTAATTAATGCAATTCGAAAAAATAATCTCCCATTTGGTG 338  
QY 1826 aagtaaaaaaacaagaaataaagacctgtgagaagaatcgttctccacatggaagac 1885  
DB 337 AAGTAAATAAAGAAATAATTAAGACCTGTGGAGAAATTCGTTGTTCCACATGGAATGGAC 278  
QY 1886 tgaactgaacgatttacaccaaaatgaacctgtctatcaagaanaatgctgcatacagta 1945  
DB 277 TGACTGTAAACGATTGACACCAAAATGACACTTGCATATAAATAATGCTCGCATACAGTA 218  
QY 1946 aactctggaatgattcttcagataatgctcctgtctttaaacaacaacattcttagagag 2005  
DB 217 AACTCTGATGATTCCTTCACATTAATGTCCTTGGCTTGACAAACACCATTTCTAGAGAG 158  
QY 2006 tctcttctcattaatcaatgaatggaatgtagtaagatgcttcttgaaaaaatgttagtc 2065

|||||  
DB 157 TCTCCTACTCATTAATTCATGAATGAATGGATTGGTAGATGCTTGTGAATAACATGTAGTC 98  
QY 2066 aagactggttaaatatcatataaagalttaacactcaatcttccaatcatataactatcc 2125  
|||||  
DB 97 AAGACAGCTGTAATAATACATATAAAGATTAAACATTCATTCATCAATCAATAACTATCC 38  
QY 2126 aaataaaaat 2135  
|||||  
DB 37 AATATAAAT 28

RESULT 8  
AM592711 544 bp mRNA 1linear EST 22-MAR-2000  
LOCUS b446606.x1 Soares\_NFL\_T\_GBC\_ST1 Homo sapiens cDNA clone  
DEFINITION IMAGE:2934946.3 similar to TR:014546 O14546 POLYSPECIFIC ORGANIC  
CATION TRANSPORTER.; mRNA sequence.  
ACCESSION AM592711  
VERSION AM592711.1 GI:7279899  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 544)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
This clone is available royalty-free through LNLN; contact the  
IMAGE Consortium (info@image.lnlnl.gov) for further information.  
Seq primer: -40UP from gibco  
High quality sequence stop: 444.  
Location/Qualifiers

1. 544  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2934946"  
/clone\_lib="Soares\_NFL\_T\_GBC\_ST1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site.1: Not I; Site.2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBH119, testis NHT, and B-cell  
NCI-CGAP GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682692-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo. "  
BASE COUNT 149 a 92 c 101 g 202 t  
ORIGIN

Query Match 23.7%; Score 505.4; DB 9; Length 544;  
Best Local Similarity 99.8%; Pred. No. 4.5e-99;  
Matches 506; Conservative 0; Mismatches 1; Indels 0;  
QY 1629 agctgactgctcgttggaatctcacaccttttccctgaaagtttggaatgact 1688  
|||||  
DB 544 AGCTGACTGCTCTGATGGAATCTTACCCCTTTTCCCTGAAAGTTTGGGAATGACT 485  
QY 1689 ctccagaaccttagagcagatgcagaagtgtaatggtltaagatcttggaataaaca 1748  
|||||  
DB 484 CTTCAGAAACCTTAGAGCAGATGAGAAAGTGAATGTTGATGATCTGGAAAAAACA 425  
QY 1749 agagactcaatggaagcagaagaatcccaaggttctaataacgcatcttgaaaaaat 1808  
|||||

Db 424 AGAGACTCAATGAGACAGAGAAAAATCCAGCTTTTAATACGCAATTCGAAAAAT 365

Qy 1809 atttaccatttggfagaagttaaaaaacgagaaaaataagacctggtgagaatttgg 1868

Db 364 ATTACCCCTTTGGTGAAGTGAAGAAACGAAAAATTAAGACCTGTGGAGAAATTCGTTG 305

Qy 1869 ttcccactgaatgagctgactgtaacgattgacaccaaaatgaacctgtcatcaagaa 1928

Db 304 TTCCCATGTAATGAGACTGACTGTAAACGATGACACCAAAATGAACCTGTCTATCAAGAA 245

Qy 1929 atgctgctataagaagtaaccttgatgatttcttccaataatgcttctgtttcaaac 1988

Db 244 ATGCTGCTACAGTAAACTCTGATGATTTCTCAAGATATATGCTTGCCTTTACAAAC 185

Qy 1989 caaccatttcaagagatcccttactcatlaattcaatgaagaatggttgaagatgctc 2048

Db 184 CACCATTTCTAGAGAGTCTCTCTACTCATTAATCAATGAAGATGATGTAAGATGTC 125

Qy 2049 ttgaatacatgttagtcaagagctggttaaatatatacatataaagathaacatctattccaa 2108

Db 124 TTGAAAAATGTTAGTCAAGAGACTGTAATAATATATTAAGATTAAACATCTCATTTCCAA 65

Qy 2109 tctataaataactatccaataaat 2135

Db 64 TCATCAAAATATCTATCAAAATAAAT 38

RESULT 9

AI628726/c 498 bp mRNA Linear EST 07-MAR-2000

LOCUS AI628726

DEFINITION ty78b05.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2285169 3'

TRANSPORTER: mRNA sequence.

ACCESSION AI628726

VERSION AI628726

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Christopher Moskalkuk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/dbp/image/image.html  
Insert Length: 595 Std Error: 0.00  
Seq primer: -40up from Gibco  
High quality sequence stop: 455  
POLYA-No.

FEATURES

SOURCE

Location/Qualifiers

1. 498

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2285169"

/clone\_lib="NCI\_CGAP\_Kid11"

/lab\_host="DH10B"

/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library

BASE COUNT 128 a 105 c 117 g 148 t

ORIGIN

Query Match 23.3%; Score 496.4; DB 9; Length 498;

Best Local Similarity 99.8%; Pred. No. 3.9e-97;

Matches 497; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1527 gtgggggtcacatccagcgcctccagatggagagatcaatggccccactttttac 1586

Db 438 GTGGGGGTCAATCCAGGCTCCAGAGTGGCAATCTTCCCTTACTTTGTTTTC 379

Qy 1587 ctgagtgcttaacaagaatgctgcccatacatgctatggttagtctgactgctgatt 1646

Db 378 CTCGGTGTTCACAAAGATGCTGCCCTACATGCGTAGTCTGACTGCTCCGAT 319

Qy 1647 ggaatctcaccctttttccctggaagtgttggaatgactcttccagaaccttaag 1706

Db 318 GGAATCTCACCTTTTTCCTGGAAGTTTGGAATGACTCTTCCAGAACTTAGAG 259

Qy 1707 cagatcgagaagtgaatggttcagatctcgggaaaaaacaagaagatcgaatgagaca 1766

Db 258 CAGATCGAGAAAGTGAATGTTTACATCTGGGAAAAAACAAGACATCATGAGACA 199

Qy 1767 gaagaatacccgaaggttctaataactgcatctgaaaaataatacccatgttgga 1826

Db 198 GAAGAAATCCCAAGGTTCTATATCATGCAATCGAAAAATATCTACCCCATTTGGTGA 139

Qy 1827 agtgaataacagaataaataaagccctgtgagaatctggttccactggaatggact 1886

Db 138 AGTGAATAACAGAAAAATTAAGACCTGTGAGAAATTCGTTGCCACTGAAATGACT 79

Qy 1887 gactgtaacgattgacaccaaaatgaacctgtctataaagaatgctgctcatcaagtaa 1946

Db 78 GACTGTAAAGATTGACACCAAAATGAACCTGTATCAGAAATGCTGTATACAGTAA 19

Qy 1947 actctgagatgattcttc 1964

Db 18 ACTGTGATGATCTTCC 1

RESULT 10

LOCUS BG723644

DEFINITION 602697810F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4829882 5'

RNA sequence.

ACCESSION BG723644

VERSION BG723644.1 GI:14002831

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHRRI), Shireki Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:



QY 614 ctcctcgtgctcgagcagctgttaacagagtttggcagaagaagcttcttcggaac 673  
 Db 590 CTCTTCGTGCTCTGGCAGCTGTGACAGAGTTGGCAGAGAAAGTCTCTTTGCTAC 649  
 QY 674 catggcgttaacagctgctcagcttcctcagatcttcacagctggagagatgt 733  
 Db 650 CATGGCTGTGACAGCTGATTCACCTTCGTGACAGATTTTTCACACCAACAGTGGAGATGT 709  
 QY 734 cactgtttatgtcactcgtt-gggacatgggcccagatctccaactatgtgttagcttca 792  
 Db 710 CACTGTGTTGTTGGCAATGTGGGGCTTGGGCCAGATCTCCAAATACGTGTG-GGCTTCA 768  
 QY 793 tactagaacagaatctctggcagatcagctgtatataatctcacaattaggatgt 852  
 Db 769 TACTAGGAACATGAA--TCTGGCAAGT--GGGACGACATATCTTCCACATTTAAGAGAGT 824  
 QY 853 gcaacttttgcagtgctatgtcgtcgcacgtgttgccttacttc 902  
 Db 825 TGACCATTTTTCACACGGCCACCATGTGCTGCTCTGTTGCTACTCTCTCC 874

RESULT 12  
 AA428395 470 bp mRNA linear EST 16-OCT-1997  
 LOCUS zva48b07.r1 Soares ovary tumor NBH07 Homo sapiens cDNA clone  
 DEFINITION IMAGE:756853 5, similar to IR:G1293672 G1293672 KIDNEY-SPECIFIC  
 TRANSPORT PROTEIN.; mRNA sequence.

ACCESSION AA428395  
 VERSION AA428395  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Martin,J., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B., White,Y., Wylie,T., Waterston,K., and Wilson,R.  
 WashU-Merck EST Project 1997

TITLE Unpublished (1997)  
 JOURNAL Contact: Wilson RK  
 COMMENT Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810

FEATURES  
 SOURCE  
 1..470  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:756853"  
 /clone\_lib="Soares ovary tumor NBH07"  
 /sex="Female"  
 /tissue\_type="ovarian tumor"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: ovary; Vector: pTV73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTCACATCTGAAGTGGAGCGCGCGGTTTTTTTTTTTTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTV73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 103 a 118 c 110 g 139 t

Query Match 21.9%; Score 466.8; DB 9; Length 470;  
 Best Local Similarity 99.6%; Pred. NO. 9.9e-91;  
 Matches 468; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1278 gccctgaatgaattccagcttaacattacagcttgcctatgtcgaagctccagg 1337  
 Db 1 GCCTGATTGAATAATCCAGCTTACATTAACAGCTGCTCTATTGCGAACCCTGCCAGG 60  
 QY 1338 cgtataataatgctcgaactactgttcttgggagagagtgctctcttaactaact 1397  
 Db 61 CGTTATATATATACCTCACTCACTGTTCTGGAGAGAGTGCTCTCTTCTTCAATCACTG 120  
 QY 1398 gtacccttgattattactcttattccattatgctgtcgtatcgtatcgtggaaattggatc 1457  
 Db 121 GTACCTGTGATTAATTAATCTTATTCATTAATGCTGCTGATCTGGGAAATTTGGGATC 180  
 QY 1458 acctctgcttctcactatgctatgtcttcaactgtcgtacgtcctaccacacctgtcagg 1517  
 Db 181 ACCTCTGCTTTCTCCATGCTGTATGCTTACAGCTGACCTCTACCCACCTGTGTCAGG 240  
 QY 1518 aacatggcgtggggtgcatccacacagcctccagagtgaggagacatctgccccctac 1577  
 Db 241 AACATGGCGGTGGGGTCCATCCACATCCAGGCTCCAGAGTGGAGCATCTTGGCCCTTAC 300  
 QY 1578 ttgtttacctcgtgtcttaacacagaatgtgccttaacatgctcagtggtgtagctgact 1637  
 Db 301 TTGTGTTACCTGGTGCTTACACAGAAATGCTCCCTACATGCTCATGGGTAGTGTGACT 360  
 QY 1638 gtccctgaatggaattctcaccctttttccctcgtgaagtgttggaatgctcttccagaa 1697  
 Db 361 GTCCCTGATTTGGAAATCCCTACCCCTTTTTCCTGAAAGTTGGGAATGCTCTCCAGAA 420  
 QY 1698 accttaagacagatgcgaagaagtgaatggttcagatctcgtggaaataaac 1747  
 Db 421 ACCTTAGACGATGCCGAAAGTGAATGTTTACATCTGCGAATAAATAC 470

RESULT 13  
 BG332869 754 bp mRNA linear EST 27-FEB-2001  
 LOCUS 602420714F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4548603 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG332869  
 VERSION BG332869.1 GI:13139307  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: c94pds-remail.nih.gov  
 Tissue Procurement: DCTD/DTF/Gazdar

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LNCM128 row: k column: 04  
 High quality sequence stop: 741.  
 Location/Qualifiers

FEATURES  
 SOURCE

1..754  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4548603"  
 /clone\_lib="NIH\_MGC\_18"  
 /tissue\_type="large cell carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Lung; Vector: pOTB7; Site\_1: XhoI; Site\_2:

ECORI: cDNA made by oligo-dT priming. Directionally cloned into EcorI/XhoI sites using the following 5' adaptor: GGCAGCAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."

BASE COUNT 111 a 253 c 249 g 141 t  
ORIGIN

Query Match 21.4%; Score 457.2; DB 10; Length 754;  
Best Local Similarity 86.9%; Pred. No. 1.4e-88;  
Matches 537; Conservative 0; Mismatches 78; Indels 3; Gaps 3;

```
QY 100 cgcgtgcatcaccgcgtgagtgcaagtttcggagcgagcgagtggaagcagtcgagcagtcg 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 GCGCTTCCCCGACCCAGCCGCGCTCTGTGGGCTGTGAGGGCGCATMGCGGACTACG 185
QY 160 acgaggtgagtcgcttcctcctgagcgagtgaggcccttcacgagcctcatctctcctgc 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 ACGAGGTGACCCGCTTCCCTGGGAGTGGGGCCCTTCACGGCCTCATCTTCCCTG 245
QY 220 tcagggcagacatcctcccaatgagcttcacatgtaatcagtcgtctcctgcgagga 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 246 TCAGGCGCAGCATMCCCAATGGCTTCACGGCCTCTCCCTGCTTCTGTGATAGCGA 305
QY 280 ccccgagacacccgctgctgagtgccgagcgccgagacatgacagcagcctgcgagcaaca 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 306 CCGCGGAGACCGGCTGCGGGGCGCGGACGCCGACACTGACAGCGCCTTGGCGACACC 365
QY 340 aagatgtccgcgtcgtcgtcggagcgagcgagtgccccacagctgcagccgctaac 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 ACACGTGTCCCACTGGCGCTGGGGACGCCGCGAGTGCCCCACAGCTGCCCGCGCTAAC 425
QY 400 ggcctgcaccatcgcccaattctcggagcgtcggagcgtgagccggggcgagcgtgagcc 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 GCGCTGCCACCATGCCCAACTTCTGGGCTTGGGCTGGAGCCGGGGCGACACTGAGACC 485
QY 460 tggggcagctggagcaggaagagctgctggaaggctggagagtcagcagcagctaac 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 486 TGGGGCACTGGAGAGAGAGACTGTCCGAGTGGCTGGAGTTCAGTCAGAGACTGTACAC 545
QY 520 tgcacacgttcgtacagcagtgatgagctggtgtgtgaggaacactggaaggtgcctca 579
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 546 TGTCCACCATTTGTGACCAAGTGAACCTGCTGTGTGAGAGAGACTGGAAG-CCCACCTCA 604
QY 580 ccaactccctgtctctcgtgaagcgctgctcctcgcgtcctcgtgtgccgggagcgtgac 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 CAATCTCTTGTCTTGTGGGTGTGCT-GTGGGCTCTTTCATTTTCAGGGCAGCTGTACG 663
QY 640 acaggttggcaggaagaagcttcttcgcaaccatgctgtacagactgagcttcagct 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 664 ACAGG-TTGGCGGAGAAAGATGCTGTGTCGACCATGAGGAGTGCAGAGAGGCTTCAGCT 722
QY 700 tccctcagagatttcctca 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 723 TCTGCAAAATCTTCTCGA 740
```

RESULT 14  
B1647860 799 bp mRNA linear EST 12-SEP-2001  
LOCUS 603276129F1 NIH\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5316559 5',  
DEFINITION mRNA sequence.  
ACCESSION B1647860  
VERSION B1647860  
KEYWORDS B1647860.1 GI:15562096  
SOURCE EST.  
ORGANISM house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 799)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9abs-r@mail.nih.gov  
Tissue Procurement: Lotmar Hennighausen Ph.D.; Chu-Xia Deng Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM1802 row: i column: 08  
High quality sequence stop: 736.

FEATURES  
source location/Qualifiers  
1..799  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5316559"  
/clone\_lib="NIH\_CGAP\_Mam3"  
/tissue\_type="tumor, gross tissue"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-Sport6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert 2 kb. Library constructed by Life  
Technologies, catalog #12017-018. Investigators providing  
samples: Lotmar Hennighausen/Chu-Xia Deng. Nature Genetics 22, 37-43  
(1999). Note: this is a NCI\_CGAP Library."

BASE COUNT 172 a 198 c 215 g 213 t  
ORIGIN 1 others

Query Match 21.1%; Score 450.6; DB 10; Length 799;  
Best Local Similarity 79.5%; Pred. No. 3.8e-87;  
Matches 582; Conservative 0; Mismatches 145; Indels 5; Gaps 4;

```
QY 369 cgcgaggtgccccacagctcagcgcgtacccgctcagcagcagcagcagcagcagcagcagc 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CGACAGGTGCTGCTAGAAATGCCCGCCCTACGACCTGCCACCATGCCCAACTTCTGTAG 60
QY 429 ctgcgagctggagcgcgggagcgagctggaacctggggcagctggaagcagagagctgctg 488
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CTAGGGCTGGAGCGGGGGCGGACCTGGACCTGGACACTGGACAGTGGAGAGCTGCCG 120
QY 489 gatggctggaggttcagcagcagcgtctacacctgtccacacctgtgaacgagtgagctcg 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GATGGCTGGAGTACGACAAAGACGCTTCTGTGTCCACCATGTGACAGAGTGGACCTG 180
QY 549 gtgtgtgagacaacatggaaggtgccccacacacctccctgtctctcgtgaagcgctgc 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GTGTGTAAAGAGATGAGAAAGCCCATCACCACCTCTTGTGTTGTTGTTGTTGTTGTTG 240
QY 609 ctgcgctccttgctgctcgggagcgtgtgacagaggttggcagaagaagcttctc 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 ATGGGCTCTTATTTACAGACAGCTCTCAGACAGG-TTGGTCGCAAGATGTGCTGTTT 299
QY 669 gcaaccatgctgtacagactgagcttcagctcctcgcagagatttctcacaagcgtggag 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 TTGACCATGGCATGACAGCTGGCTTCACGTTCTCTGCAAGTCTTCTGTGAACCTTCAG 359
QY 729 atgtcaactgttatttgcatactgtggcagtgagcagcagatcccaacatgctgtagc 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 360 AATTTACAGTGTCTTTTGTCTGTTGNGCATGGGTCAGATGTCCACAGTGTGACACCA 419
QY 789 ttcatactaggaagaagaattcttggcaagtcagtcagtcgtatataatcttcacattaga 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 TTTGTCTGGGAACAAATTTCTTTCACATTAATTAATTAATTAATTAATTAATTAATTA 479
QY 849 gtgtgcacatttctgcagttgctatagctgcgtgcgcacctg-ttggcttactcaatag 907
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 GT-TGCATTTTATGCGTTTGCTTCATGGTGTGCTGACACTGTCTTTGTCATCTTCATCAG 538
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 02:03:06 ; Search time 86.06 Seconds  
(without alignments)  
718.895 Million cell updates/sec

Title: US-09-521-195B-3

Perfect score: 2883  
Sequence: 1 MRDYDEVTAFLGEMWPPQRL.....HTFMKDGQERPTLKSTAF 557

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

A.Geneseq\_032802:\*  
1: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SID55/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SID55/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SID55/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2883	100.0	557	20	AAV01650
2	2883	100.0	557	21	AAV83929
3	2883	100.0	557	22	ABG03029
4	2513	87.2	557	20	AAV01652
5	2513	87.2	557	21	AAV83930
6	2513	87.2	557	21	AAV83930
7	2295.5	79.6	557	21	AAV01649
8	2218	74.5	553	20	AAV01651
9	2148	74.5	553	21	AAV01651
10	2148	74.5	553	21	AAV01651
11	885	30.7	548	22	ABB62279

12	816	28.3	567	22	ABB68354
13	802.5	27.8	577	22	AAV78357
14	785.5	27.2	560	21	AAV43038
15	784	27.2	584	22	AAV00930
16	753.5	26.1	561	22	ABB60830
17	753.5	26.1	585	22	AAV79351
18	738.5	25.6	565	22	AAV11289
19	738.5	25.6	565	22	ABG11291
20	735	25.5	554	19	AAV64538
21	727	25.2	555	22	AAV79251
22	724.5	25.1	540	22	AAV49401
23	719	24.9	535	21	AAV51249
24	716.5	24.9	674	22	ABB58206
25	712	24.7	556	17	AAV77676
26	701.5	24.3	538	22	AAV47272
27	697	24.2	546	22	AAV47272
28	693	24.0	542	21	AAV92802
29	690	23.9	548	21	AAV08823
30	690	23.9	552	22	ABB12030
31	689	23.9	542	22	AAV47274
32	686.5	23.8	551	21	AAV08824
33	675.5	23.4	553	22	AAE10332
34	673	23.3	561	18	AAV44196
35	668	23.2	578	22	AAE06571
36	660.5	22.9	607	21	AAH12131
37	656.5	22.8	536	21	AAV92903
38	652.5	22.6	537	18	AAV44195
39	651	22.6	557	22	ABB71470
40	650	22.5	538	22	ABB60174
41	644	22.3	545	22	ABB36553
42	641	22.2	550	21	AAV44278
43	641	22.2	550	21	AAV47271
44	638.5	22.1	551	20	AAV84488
45	638.5	22.1	762	22	ABB63300

## ALIGNMENTS

RESULT 1	AAV01650	AAV01650 standard; Protein: 557 AA.
ID	AAV01650;	
XX	AAV01650;	
AC	AAV01650;	
XX	AAV01650;	
DT	23-JUN-1999 (first entry)	
XX		
DE	A protein with cation transporting activity.	
XX		
KW	Organic cation transporter; OCT1; OCT2; drug development; fatty liver;	
KW	heart disease; cancer; anti-tumour drug; anticancer drug.	
XX		
OS	Homo sapiens.	
XX		
PN	W09913072-A1.	
XX		
PD	18-MAR-1999.	
XX		
PF	07-SEP-1998; 98WO-JP04009.	
XX		
PR	20-MAY-1998; 98JP-0156660.	
PR	08-SEP-1997; 97JP-0260972.	
XX		
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.	
XX		
PI	Nezu J, Oku A;	
XX		
XX	WPI, 1999-215062/18.	
DR	N-PSDB; AAX26880.	
XX		
PT	Genes homologous with organic cation transporters OCT1 and OCT2,	
PT	useful in design of new drugs for treatment of diseases due to	
PT	abnormality of the transporter functions	
PT		

Drosophila melanog  
Human protein SEQ  
Human ORFX ORF2802  
Human bone marrow  
Drosophila melanog  
Human protein SEQ  
Novel human diagno  
Novel human diagno  
Human liver cell c  
Human protein SEQ  
Murine organic an  
Rat liver anion tr  
Drosophila melanog  
Rat OCT-1 protein.  
HOAT2B. Homo sapi  
HOAT2A. Homo sapi  
Human cerebral org  
A human organic an  
Human liver-specific  
HOAT3. Homo sapi  
A human organic an  
Human osteoclast t  
Human osteoclast t  
Human protein havi  
Hydrophobic domain  
Rat cerebral organ  
Mouse osteoclast t  
Drosophila melanog  
Mouse organic anio  
Human organic anio  
HOAT1. Homo sapi  
Rat organic anion  
Drosophila melanog

XX Claim 1; Page 51-55; 97pp; Japanese.  
PS  
XX  
CC The present sequence represents a protein with cation transporting  
CC activity. The genes are significantly homologous with organic cation  
CC transporters OCT1 and OCT2. The genes may used in drug development,  
CC particularly in the treatment of diseases due to abnormality of the  
CC organic cation transporter functions e.g. fatty liver, heart diseases  
CC and cancers, by controlling such as by inhibition or activation.  
CC Administration of anti-tumour and anticancer drugs in combination with  
CC a transporter protein inhibiting agent allow the agents to penetrate  
CC into the diseased cells to enhance the drug action.  
XX  
SQ Sequence 557 AA:  
  
Query Match 100.0%; Score 2883; DB 20; Length 557;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MDDYEVTAFLGEMGFORLIFELLSASTIPNGFTGLSVFLATPEHRCVPDAANLSS 60  
DB 1 mdydevtaflgwgpfqrlifllsasllpnqftglsvflatpchrpdaanlss 60  
XX  
XX  
QY 61 AMRNHTVPLRLDGRVPHSCRRYRLATIANFSALEPGRDVDLQLEBESCLDGMERS 120  
DB 61 awrnhtvplrltdgrevpshscrrylatiansafalepgrdvdqlqlescldgwets 120  
XX  
QY 121 QDYLSTIVTEMNLVCEDDMKAPLTISLFYGVLLSGFISGQLSDRGRKNVLFVTMGQ 180  
DB 121 qdylstivtemnlvceddmkapltislfygvllsgfisdgrgrknvlfvtmgq 180  
XX  
QY 181 TGFSEFLQIFSKNEMFVLFVLVGMGOISNVAAFEVLTGELKSVRIIFSTLGVCIPTA 240  
DB 181 tgfseflqifsknfemfvlfvlvymgoisnyaafevltgclsksvriifstlgyvcifya 240  
XX  
QY 241 FGYMVLPLFAFTRIDWRMLVALTMPGVLCVALLMWFPESSRMLISQGRFEAEVIIRKA 300  
DB 241 fgymvplpfayftridwrmlvaltmpgvlcvalmwfpesprwlisqgrfeaevliirka 300  
XX  
QY 301 AKANGIVPSTIFDSELDSSKKQOSHNLIDLRTWNIIMWTISIMLMWTISVGYFG 360  
DB 301 akangivpstifdpseldsskkqshnldlrtwnimwtisimlmwtisvgyfg 360  
XX  
QY 361 LSLDTPNLHGDI FVNCFTISAMVEVPAYVLA MLQYPRRYSMAALFLGSVLLFMQLV 420  
DB 361 lslatpnlhgdlfvncftisamvevpayvllawllqyprysmataalfgsvllfmqlv 420  
XX  
QY 421 PPDLYLATVLMVGKFGVTAFAFMSVYVYTAELYPVVRMVGVSSTASRLISPYF 480  
DB 421 ppdlylatvlmvgkfgytaafsmvyvyytaelyptvvrmmgvsvstasrlispyf 480  
XX  
QY 481 VYLGADRLPLPYILMGSITJLTALITLFLPESFGTLPDITIDMLRVKMKHKKTPSHNR 540  
DB 481 vylgadrlplpyilmgstjltaltlflpesfgtprdtidmrlvkmkhrktpshnr 540  
XX  
QY 541 MKDGOERPTIKSTAF 557  
DB 541 mkdgoerptikstaf 557  
XX  
RESULT 2  
AAV83929  
ID AAV83929 standard; Protein: 557 AA.  
AC AAV83929;  
XX  
XX 05-JUN-2000 (first entry)  
DT  
XX  
DE Human carnitine transporter protein OCTN2.  
XX  
XX Organic cation transportation; human; carnitine transporter protein;  
KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;

KW juvenile visceral steatosis.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200014210-A1.  
PN  
XX  
XX 16-MAR-2000.  
PD  
XX  
XX 07-SEP-1999; 99WO-JP04853.  
PF  
XX  
XX 07-SEP-1998; 98JP-0252683.  
PR  
XX  
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
PA  
XX  
XX Nezu J, Oku A;  
PI  
XX  
XX WPI: 2000-256966/22.  
DR  
XX  
XX N-PSDB: AAA09889.  
PT  
XX  
XX Systemic carnitine deficiency gene OCTN2 encoding part of organic  
PT cation transporter, useful as diagnostic tool -  
XX  
XX Example 1; Page 36-40; 106pp; Japanese.  
XX  
CC This sequence represents the human carnitine transporter protein OCTN2.  
CC The coding sequence can be used as a target for diagnosis of systemic  
CC carnitine deficiency by detecting the presence of mutations in the  
CC sequence, especially seen in the disease juvenile visceral steatosis  
CC (jvs). The wild type OCTN2 gene can be used in the gene therapy of the  
CC disease state.  
XX  
SQ Sequence 557 AA:  
  
Query Match 100.0%; Score 2883; DB 21; Length 557;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MRDYDEVTAFLGEMGFORLIFELLSASTIPNGFTGLSVFLATPEHRCVPDAANLSS 60  
DB 1 mrdydevtaflgwgpfqrlifllsasllpnqftglsvflatpchrpdaanlss 60  
XX  
QY 61 AMRNHTVPLRLDGRVPHSCRRYRLATIANFSALEPGRDVDLQLEBESCLDGMERS 120  
DB 61 awrnhtvplrltdgrevpshscrrylatiansafalepgrdvdqlqlescldgwets 120  
XX  
QY 121 QDYLSTIVTEMNLVCEDDMKAPLTISLFYGVLLSGFISGQLSDRGRKNVLFVTMGQ 180  
DB 121 qdylstivtemnlvceddmkapltislfygvllsgfisdgrgrknvlfvtmgq 180  
XX  
QY 181 TGFSEFLQIFSKNEMFVLFVLVGMGOISNVAAFEVLTGELKSVRIIFSTLGVCIPTA 240  
DB 181 tgfseflqifsknfemfvlfvlvymgoisnyaafevltgclsksvriifstlgyvcifya 240  
XX  
QY 241 FGYMVLPLFAFTRIDWRMLVALTMPGVLCVALLMWFPESSRMLISQGRFEAEVIIRKA 300  
DB 241 fgymvplpfayftridwrmlvaltmpgvlcvalmwfpesprwlisqgrfeaevliirka 300  
XX  
QY 301 AKANGIVPSTIFDSELDSSKKQOSHNLIDLRTWNIIMWTISIMLMWTISVGYFG 360  
DB 301 akangivpstifdpseldsskkqshnldlrtwnimwtisimlmwtisvgyfg 360  
XX  
QY 361 LSLDTPNLHGDI FVNCFTISAMVEVPAYVLA MLQYPRRYSMAALFLGSVLLFMQLV 420  
DB 361 lslatpnlhgdlfvncftisamvevpayvllawllqyprysmataalfgsvllfmqlv 420  
XX  
QY 421 PPDLYLATVLMVGKFGVTAFAFMSVYVYTAELYPVVRMVGVSSTASRLISPYF 480  
DB 421 ppdlylatvlmvgkfgytaafsmvyvyytaelyptvvrmmgvsvstasrlispyf 480  
XX  
QY 481 VYLGADRLPLPYILMGSITJLTALITLFLPESFGTLPDITIDMLRVKMKHKKTPSHNR 540  
DB 481 vylgadrlplpyilmgstjltaltlflpesfgtprdtidmrlvkmkhrktpshnr 540  
XX

OY 541 MLKDGERTILKSTAF 557  
 ||||||||||||||||  
 Db 541 mlkdgertilkstaf 557

# RESULT 3

ABG03029  
 ID ABG03029 standard; Protein; 557 AA.

AC ABG03029;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #3020.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YF;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS67216.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

PS Claim 20: SEQ ID NO 33388; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG030377 represent novel human

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC  
 XX  
 SQ Sequence 557 AA;

Query Match 100.0%; Score 2883; DB 22; Length 557;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRDYDEVTAFLGEGPFOBLIFELLASAIIPNGFTGLASVFELIATPEHRCVPPAANLSS 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 mrdydevtaflgwegpfgqrlifllsasliipngftglasvfliatpehrcrvpdaanls 60  
 OY 61 AMRNHTVPLRLRDGREGVPHSCRRYRLATTANFSALGLEPGRVDVGLQEGESCLDGEFES 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 awrnhtvplrlrdgrevpshscrrylattantfsalglepgrvdiqlgescdlgyefs 120  
 OY 121 ODVYLSITVTEENMLVCEDDMKAPLTSLFFVGVLLGSFISGOLSPGRKNVLPFTMGMO 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 121 qdvylsitvtemlvceddmkapltslffvylvllgstisgqlsdrfgknylvtummg 180  
 OY 181 TGSEFLQIFSKNPEMFVFLVVGMOISNYAAEVLGTEILGKSVRIIFSTLGVCIPIYA 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 tgsflqifsknfemfvlvflvgmqisnyaaefylgteilgkavrllfstlgyclfya 240  
 OY 241 FGYNVPLPFAVFRDMRMMLVALTMPGVLCVALMWFIPSPRWLISOGREFEAEYIRKA 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 241 fgyinvplfayfrdmrmmlvaltmprvllcvalmwfipspwllisogrefeaevalrka 300  
 OY 301 AKANGIVVPSTIPDSELDLSSKKQOSHNIIDLRTWNI RMVVTIMSI MLMTISVGYG 360  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 301 akangivvpstifdpseldlsskkgshnidlirtwnirmvvtimslmlmtisvgyfg 360  
 OY 361 LSLDTPNLHGDIIVNCPFLSAMVEVPAYVLAMLLIQLPRRYSMATPLFGSVLLFMOLV 420  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 361 lsltdpnlhgdiivncpflsamvevpayvllmllqylprrysmatplfgsvllfmolv 420  
 OY 421 PPDIYVLTATVVMVKGEGVAAFSMYVYTAELPYVAVNMKGVSSTASRGLSTLSYF 480  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 421 pldiyvltatvwmvkgfgyaaafsmvyytaelpyvavnmkgvssstasrlstlsyft 480  
 OY 481 VYLGAVDRFLPYLIMSGSLFLLTAIILTLFLPESFGPPLPTIDOMLRVKGMRKTPSTR 540  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 481 vylgavdrflpylimsgslflltaliiltflpessfgpplptidomlrvgmrktrpstr 540  
 OY 541 MLKDGERTILKSTAF 557  
 ||||||||||||||||  
 Db 541 mlkdgertilkstaf 557

## RESULT 4

ID AAY01652 standard; Protein; 557 AA.

AC AAY01652;

DT 23-JUN-1999 (first entry)

DE A protein with cation transporting activity.

KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;  
 heart disease; cancer; anti-tumour drug; anticancer drug.

OS Mus musculus.

PN W09913072-A1.

PD 18-MAR-1999.

PF 07-SEP-1998; 98WO-JP04009.

PR 20-MAY-1998; 98JP-0156660.

PR 08-SEP-1997; 97JP-0260972.

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

PI Nezu J, Oku A;

DR WPI: 1999-215062/18.

DR N-PSDB; AAX26902.

PT Genes homologous with organic cation transporters OCT1 and OCT2,

PT useful in design of new drugs for treatment of diseases due to  
 PT abnormality of the transporter functions

PS Claim 1: Page 75-79; 97pp; Japanese.

CC The present sequence represents a protein with cation transporting  
 CC activity. The genes are significantly homologous with organic cation  
 CC transporters OCT1 and OCT2. The genes may used in drug development,  
 CC particularly in the treatment of diseases due to abnormality of the  
 CC organic cation transporter functions e.g. fatty liver, heart diseases  
 CC and cancers, by controlling such as by inhibition or activation.  
 CC Administration of anti-tumour and anticancer drugs in combination with  
 CC a transporter protein inhibiting agent allows the agents to penetrate  
 CC into the diseased cells to enhance the drug action.

XX Sequence 557 AA:

Query Match 87.2%; Score 2513; DB 20; Length 557;  
 Best Local Similarity 85.5%; Pred. No. 5e-264;  
 Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEMGPORLIFELLSASTIIPNGFTGLSSVFELIATPEHRCRYDPAANLSS 60  
 DB 1 mrdydevtaflgempqrllifllsaalipngfngmsivflagpcheiclyphvniss 60  
 QY 61 AMRNHTVPLRLRDGREVPHSCRRYRLATTIANFSALGLEPGRVDVLGQLEOESCLDGMERS 120  
 DB 61 awrnhsipletkdgrvqpkcrryrlattianfselglepgrvdleqlegescldgweyd 120  
 QY 121 QDYVLSTIVTEWNLVCEDDMKAPLITSLFVGVLLGSFISGOLSPFGKKNVLFVTMGQ 180  
 DB 121 kdyvlstivtewlvckddmkapltslfvgvllgmsfissgqlstfgrkknvlfvtmgq 180  
 QY 181 TGFSEFLQIFSKNEMFVFLVGMGOISNYAAFLVLTGLTEILGKSVRIIFSTLGVCIFFYA 240  
 DB 181 tgfseflqifsknemfvlvlgmgoisnyaaflvltglteilgksvriifstlgyvcifya 240  
 QY 241 FGYWVLPFLFAYFTRDMRMLVALTMPGVLCVALMWFIPESPRLISQGFEEAEVITRKA 300  
 DB 241 fgfwnvlpflfayftrdmrmlvaltmpgvlcvalmwfipesprrllsqgfikeaevilitrka 300  
 QY 301 AKANGIVPSTIPDPELQDLSKKOOSHNIIDLRTWNIIRMTYIMSIMLMTISVGYEG 360  
 DB 301 akngivpstifdpeseldlsskkooshnidlrrtwniirmtymisimlmtisvgyeg 360  
 QY 361 LSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLLIQYLPRTYSMAATALFLGGSVLLFMQ 420  
 DB 361 lsldtpnlhgdifvncflsamvevpayvllawlllqylprtyssaalfggsvllfmqlv 420  
 QY 421 PRDIYIATVIVMGKFGVTAASMYVYVAELVPTVVRNMGVGVSTASRLGISPYE 480  
 DB 421 prdiyatvivmgkfgvtaasmvvyvaelvptvvrnmvgvstasrlgislspye 480  
 QY 481 VYLAVYDFELIYIMGSLTILTAITLFLPESEFGTPPLDITDQMLRVKGMKHKRTPSHTR 540  
 DB 481 vylavdydfelilyimgsltltailtflfpefsfgvtpdtdqmlrvkgkqkwgsgqtr 540  
 QY 541 MLKDGQERPTILKSTAF 557  
 DB 541 mqkdgeesptvlkstaf 557

RESULT 5

ID AAB20580 standard; Protein; 557 AA.

XX AAB20580;

XX 11-DEC-2000 (first entry)

XX Mouse OCTN2 amino acid sequence.

KW Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;  
 KW identification; regulator; carnitine transport.

OS Mus musculus.

PN WO200046368-A1.

PD 10-AUG-2000.

XX 04-FEB-2000; 2000MO-JP00619.

XX 05-FEB-1999; 99JP-0028406.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Nezu J, Ose A;

XX WPI, 2000-586982/55.

PT Organic cation transporter gene OCTN3 expressed in testis for  
 PT identification of regulators of carnitine transport for use as drugs  
 PS Example 2; Fig 2; 58pp; Japanese.

CC The present invention describes a mouse organic cation transporter  
 CC protein (OCTN3). Also described are: (1) a method for screening  
 CC compounds for their ability to regulate the transport of an organic  
 CC cation into the cell, by generating a cell expressing OCTN3 at the  
 CC cell membrane, contacting with the compound and organic cation, and  
 CC observing the degree of transport of the organic cation; and (2) a  
 CC method for screening compounds for their ability to be transported into  
 CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell  
 CC membrane, contacting with the compound and observing the degree of  
 CC transport of the compound. OCTN3 can be used for the identification of  
 CC regulators of the transport of organic cations (especially carnitine)  
 CC into cells by OCTN3, for use as drugs. The present sequence represents  
 CC the mouse OCTN2 amino acid sequence, which is used in an example from  
 CC the present invention.

XX Sequence 557 AA:

Query Match 87.2%; Score 2513; DB 21; Length 557;  
 Best Local Similarity 85.5%; Pred. No. 5e-264;  
 Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;

QY 1 MRDYDEVTAFLGEMGPORLIFELLSASTIIPNGFTGLSSVFELIATPEHRCRYDPAANLSS 60  
 DB 1 mrdydevtaflgempqrllifllsaalipngfngmsivflagpcheiclyphvniss 60  
 QY 61 AMRNHTVPLRLRDGREVPHSCRRYRLATTIANFSALGLEPGRVDVLGQLEOESCLDGMERS 120  
 DB 61 awrnhsipletkdgrvqpkcrryrlattianfselglepgrvdleqlegescldgweyd 120  
 QY 121 QDYVLSTIVTEWNLVCEDDMKAPLITSLFVGVLLGSFISGOLSPFGKKNVLFVTMGQ 180  
 DB 121 kdyvlstivtewlvckddmkapltslfvgvllgmsfissgqlstfgrkknvlfvtmgq 180  
 QY 181 TGFSEFLQIFSKNEMFVFLVGMGOISNYAAFLVLTGLTEILGKSVRIIFSTLGVCIFFYA 240  
 DB 181 tgfseflqifsknemfvlvlgmgoisnyaaflvltglteilgksvriifstlgyvcifya 240  
 QY 241 FGYWVLPFLFAYFTRDMRMLVALTMPGVLCVALMWFIPESPRLISQGFEEAEVITRKA 300  
 DB 241 fgfwnvlpflfayftrdmrmlvaltmpgvlcvalmwfipesprrllsqgfikeaevilitrka 300  
 QY 301 AKANGIVPSTIPDPELQDLSKKOOSHNIIDLRTWNIIRMTYIMSIMLMTISVGYEG 360  
 DB 301 akngivpstifdpeseldlsskkooshnidlrrtwniirmtymisimlmtisvgyeg 360  
 QY 361 LSLDTPNLHGDIIVNCFLSAMVEVPAYVLAWLLIQYLPRTYSMAATALFLGGSVLLFMQ 420  
 DB 361 lsldtpnlhgdifvncflsamvevpayvllawlllqylprtyssaalfggsvllfmqlv 420

```

QY 421 PPDLVYLATVLMVNGKFGVTAAFSMVYVYTAELPTVNMGVGVSSTASRLGSIISPYF 480
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 421 pseelfystalvmvngkfgltsaymvyvyaelyptvnmgyvvsstasrlgslispyf 480
QY 481 VYLGAYDRFLPYILMGSLITLITLFLPESFGTPLPPTIDOMLRVKGMRKRTPSHTR 540
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 481 vylgaydrflpyilmgslitliltlflfespfgvplpdtldqmlrvkgikqwgqigsqr 540
QY 541 MLKDGERPTILKSTAF 557
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 541 mskdgesptvlkstaf 557

RESULT 6
AA83930
ID AA83930 standard; Protein; 557 AA.
XX
AC AA83930;
XX
DT 05-JUL-2000 (first entry)
XX
DE Mouse carnitine transporter protein OCTN2.
XX
KW Organic cation transportation; mouse; carnitine transporter protein;
KW OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;
KW juvenile visceral steatosis.
XX
OS Mus musculus.
XX
PN WO200014210-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-JP04853.
XX
PR 07-SEP-1998; 98JP-0252683.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI Nezu J, Oku A;
XX
DR WPI: 2000-256966/22.
DR N-PSDB; AAA09890.
XX
PT Systemic carnitine deficiency gene OCTN2 encoding part of organic
PT cation transporter, useful as diagnostic tool
XX
PS Example 1; Page 47-51; 106pp; Japanese.
XX
CC This sequence represents the mouse carnitine transporter protein OCTN2.
CC The coding sequence of the corresponding human protein can be used as a
CC target for diagnosis of systemic carnitine deficiency by detecting the
CC presence of mutations in the sequence, especially seen in the disease
CC juvenile visceral steatosis (jvs). The wild type OCTN2 gene can be used
CC in the gene therapy of the disease state.
XX
SQ Sequence 557 AA;

Query Match 87.2%; Score 2513; DB 21; Length 557;
Best Local Similarity 85.5%; Pred. No. 5e-264;
Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;
QY 1 MRDYEVTAFIGWGFQRIIFLLASIIINGFTGSSVFLIADPEHRCVPDANLSS 60
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1 mrdyevtafigwgfqriifllasiiingftgssvfladpehrcvpdanlss 60
QY 61 AMRNTVPLRLDRGEVPHSCRRIRLATIANFSAIGLEPGHVDVLGOLQESCLDGWEFS 120
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 61 amrntvplrldrgevpshscrrirlatianf saiglepg hvdvlgolqescldgweyfd 120
QY 121 QDVYLTIVTEWNLVCEDDWKAPLITSLFFYGVLLGSTRISQSLDRGRKNVLEVTMGMO 180
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

```

Db 121 kdvfslvtewdlvckddwkaplitslffvgvlmgstfsglsdrigrknvflitmgmq 180
QY 181 TGFSLQIFSKNFEMEFVVLVVLVGMGOISNYVAFLGTEILGKSVRIIFSTLGVCFYA 240
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 181 tgfslqyfsknfemefvvlvvlvlgmgqisnyvaafvgtelgksvriifstlgvcfyfa 240
QY 241 FGWMLPLFAFIDMRRLVALLMPGVLCAVLMWFTIPESRWLISGREGAEAVITRKA 300
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 241 fgwmlplfayfildmrllalalpvgvlcgalwffipesprwlisgrikaeavilrka 300
QY 301 AKANGIVPSTIPDPSELQDSSKKOOSHNTIDLRTMNIRMTIMSGMIMMTISGVFG 360
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 301 akingivapstifdpseiqdlnstckpqhhlydlitcrnlvltmslilwltisvayig 360
QY 361 LSLDTPNLHSDIYNCLFASAVVEPAAVLANLLQYLPRRYSMAATALFLGSVLLFMOVLV 420
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 361 lsldtpnlhgdlyncfllaavepaylawlllqypryslsaafflgsvllfmqlv 420
QY 421 PPDLVYLATVLMVNGKFGVTAAFSMVYVYTAELPTVNMGVGVSSTASRLGSIISPYF 480
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 421 pseelfystalvmvngkfgltsaymvyvyaelyptvnmgyvvsstasrlgslispyf 480
QY 481 VYLGAYDRFLPYILMGSLITLITLFLPESFGTPLPPTIDOMLRVKGMRKRTPSHTR 540
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 481 vylgaydrflpyilmgslitliltlflfespfgvplpdtldqmlrvkgikqwgqigsqr 540
QY 541 MLKDGERPTILKSTAF 557
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 541 mskdgesptvlkstaf 557

```

```

RESULT 7
AAB20578
ID AAB20578 standard; Protein; 564 AA.
XX
AC AAB20578;
XX
DT 11-DEC-2000 (first entry)
XX
DE Mouse OCTN3 protein SEQ ID NO:1.
XX
KW Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
KW identification; regulator; carnitine transport.
XX
OS Mus musculus.
XX
PN WO200046368-A1.
XX
PD 10-AUG-2000.
XX
PF 04-FEB-2000; 2000MO-JP00619.
XX
PR 05-FEB-1999; 99JP-0028406.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI Nezu J, Ose A;
XX
DR WPI: 2000-586982/55.
DR N-PSDB; AAA88053.
XX
PT Organic cation transporter gene OCTN3 expressed in testis for
PT identification of regulators of carnitine transport for use as drugs
XX
PS Claim 1; Page 34-39; 58pp; Japanese.
XX
CC The present invention describes a mouse organic cation transporter
CC protein (OCTN3). Also described are: (1) a method for screening
CC compounds for their ability to regulate the transport of an organic
CC cation into the cell, by generating a cell expressing OCTN3 at the
CC cell membrane, contacting with the compound and organic cation, and
CC observing the degree of transport of the organic cation; and (2) a

```

CC method for screening compounds for their ability to be transported into  
CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell  
CC membrane, contacting with the compound and observing the degree of  
CC transport of the compound. OCTN3 can be used for the identification of  
CC regulators of the transport of organic cations (especially carnitine)  
CC into cells by OCTN3, for use as drugs. The present sequence represents  
CC mouse OCTN3.

XX Sequence 564 AA;

Query Match 79.6%; Score 2295.5; DB 21; Length 564;  
Best Local Similarity 80.6%; Pred. No. 2,3e-240;  
Matches 435; Conservative 54; Mismatches 48; Indels 3; Gaps 1;

QY 1 MRDYEVATFGEMGPPRRLIFFLASIIIPNGFTGLSSVFLIATPEHRCRPDANLSS 60  
DB 1 mldyevatafigewgtfrllifllssailpnngftglssavflitapehrcrtpdtnlss 60  
QY 61 AMRNHTVPLRLDGRREVPHSCRRYRLATIANFSALEPGSDVDLGQLEQESCLDGEFES 120  
DB 61 awrnhtvplrlldgrrevphscrryrlatianfselepgsdvdleqesclidgewyd 120  
QY 121 ODVYLSITVTENMLVCEEDMKAPLTISLFEVGVLLGSFISQLSDRGRKNVLFVTGMQ 180  
DB 121 kdlfistltvewdlvckdawkapltisfeyvllgsfisdrgknvlfvtgmah 180  
QY 181 TGFSFLQIFSKNEMFVVLVYVGMGOISNYVAFAVLGTEILGKSVRIIFSTLGCIFYA 240  
DB 181 tgfslfqlfksknemfvlvlyvngmgoisnyvaafvlgteilgksvriifstlgcifya 240  
QY 241 FGYWVLPFAFTRIRRMRLVALTMPGVLCAVLMWTFEPSPRWLISQGREAEAVIIRKA 300  
DB 241 fgywvlpfaftrirrmrlvaltmpgvlcalvmwtfepspRWLISQGREAEAVIIRKA 300  
QY 241 fgywvlpfaftrirrmrlvaltmpgvlcalvmwtfepspRWLISQGREAEAVIIRKA 300  
DB 241 fgywvlpfaftrirrmrlvaltmpgvlcalvmwtfepspRWLISQGREAEAVIIRKA 300  
QY 301 AKANGIVPSTIFDPSE--LQDLSKKQOOSHNIIDLRTNIRMTIMSLMWTISVIG 357  
DB 301 akangivpstifdpse--lqdlsskkqooshnidlRTNIRMTIMSLMWTISVIG 357  
QY 360 AKANGIVPSTIFDPSE--LQDLSKKQOOSHNIIDLRTNIRMTIMSLMWTISVIG 360  
DB 360 akangivpstifdpse--lqdlsskkqooshnidlRTNIRMTIMSLMWTISVIG 360  
QY 361 YFGLSLDPNNLHGDIFVNCFLSAMVEPAVYLAWLLOYPRRYSMATALFGSVLLFM 417  
DB 361 yfglslDPNNLHGDIFVNCFLSAMVEPAVYLAWLLOYPRRYSMATALFGSVLLFM 417  
QY 418 OLVPDXYLATVLMWKGKGTAFSAFVYVYTAELPTVVRNMGVGSSTASRLGSLIS 477  
DB 418 olvpdxyLATVLMWKGKGTAFSAFVYVYTAELPTVVRNMGVGSSTASRLGSLIS 477  
QY 421 GLVPEDLHYISCTLVWVGKFGITLSAYSMVYVYTAELPTVVRNMGVGSSTASRLGSLIS 480  
DB 421 glvpEDLHYISCTLVWVGKFGITLSAYSMVYVYTAELPTVVRNMGVGSSTASRLGSLIS 480  
QY 478 PFYVVLGADRFPLPYILMGSLTILTAITLFLPESGTPLPDITDQMLRVKGMKHKRTPS 537  
DB 478 pfyvVLGADRFPLPYILMGSLTILTAITLFLPESGTPLPDITDQMLRVKGMKHKRTPS 537  
DB 481 PFYVYLGYAYDRPLPYILMGSLTILTAITLFLPESGTPLPDITDQMLRVKGMKHKRTPS 540

RESULT 8

AA01649  
ID AA01649 standard; Protein: 551 AA.

AA01649;  
AC AA01649;

DT 23-JUN-1999 (first entry)

DE A protein with cation transporting activity.

KM Organic cation transporter; OCT1; OCT2; drug development; fatty liver;  
KM heart disease; cancer; anti-tumour drug; anticancer drug.

OS Homo sapiens.

PN WO913072-A1.

XX 18-MAR-1999.

PF 07-SEP-1998; 98WO-JP04009.

PR 20-MAY-1998; 98JP-0156660.

PR 08-SEP-1997; 97JP-0260972.

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

PI Nezu J, Oku A;

DR WPI, 1999-215062/18.

DR N-PSDB; AAX26879.

PT Genes homologous with organic cation transporters OCT1 and OCT2,

PT useful in design of new drugs for treatment of diseases due to

PT abnormality of the transporter functions

PS Claim 1; Page 41-45; 97pp; Japanese.

CC The present sequence represents a protein with cation transporting

CC activity. The genes are significantly homologous with organic cation

CC transporters OCT1 and OCT2. The genes may be used in drug development,

CC particularly in the treatment of diseases due to abnormality of the

CC organic cation transporter functions e.g. fatty liver, heart diseases

CC and cancers, by controlling such as by inhibition or activation.

CC Administration of anti-tumour and anticancer drugs in combination with

CC a transporter protein inhibiting agent allows the agents to penetrate

CC into the diseased cells to enhance the drug action.

CC Sequence 551 AA;

Query Match 76.9%; Score 2218; DB 20; Length 551;  
Best Local Similarity 75.9%; Pred. No. 5.9e-232;  
Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

QY 1 MRDYEVATFGEMGPPRRLIFFLASIIIPNGFTGLSSVFLIATPEHRCRPDANLSS 60  
DB 1 mrdyevatafigewgtfrllifllssailpnngftglssavflitapehrcrtpdtnlss 60  
QY 61 AMRNHTVPLRLDGRREVPHSCRRYRLATIANFSALEPGSDVDLGQLEQESCLDGEFES 120  
DB 61 awrnhtvplrlldgrrevphscrryrlatianfselepgsdvdleqesclidgewyd 120  
QY 121 ODVYLSITVTENMLVCEEDMKAPLTISLFEVGVLLGSFISQLSDRGRKNVLFVTGMQ 180  
DB 121 qdvylstvtvewdlvckdawkapltisfeyvllgsfisdrgknvlfvtgmah 180  
QY 181 TGFSFLQIFSKNEMFVVLVYVGMGOISNYVAFAVLGTEILGKSVRIIFSTLGCIFYA 240  
DB 181 tgfslfqlfksknemfvlvlyvngmgoisnyvaafvlgteilgksvriifstlgcifya 240  
QY 241 FGYWVLPFAFTRIRRMRLVALTMPGVLCAVLMWTFEPSPRWLISQGREAEAVIIRKA 300  
DB 241 fgywvlpfaftrirrmrlvaltmpgvlcalvmwtfepspRWLISQGREAEAVIIRKA 300  
QY 241 fgywvlpfaftrirrmrlvaltmpgvlcalvmwtfepspRWLISQGREAEAVIIRKA 300  
DB 241 fgywvlpfaftrirrmrlvaltmpgvlcalvmwtfepspRWLISQGREAEAVIIRKA 300  
QY 301 AKANGIVPSTIFDPSE--LQDLSKKQOOSHNIIDLRTNIRMTIMSLMWTISVIG 357  
DB 301 akangivpstifdpse--lqdlsskkqooshnidlRTNIRMTIMSLMWTISVIG 357  
QY 360 AKANGIVPSTIFDPSE--LQDLSKKQOOSHNIIDLRTNIRMTIMSLMWTISVIG 360  
DB 360 akangivpstifdpse--lqdlsskkqooshnidlRTNIRMTIMSLMWTISVIG 360  
QY 361 YFGLSLDPNNLHGDIFVNCFLSAMVEPAVYLAWLLOYPRRYSMATALFGSVLLFM 417  
DB 361 yfglslDPNNLHGDIFVNCFLSAMVEPAVYLAWLLOYPRRYSMATALFGSVLLFM 417  
QY 418 OLVPDXYLATVLMWKGKGTAFSAFVYVYTAELPTVVRNMGVGSSTASRLGSLIS 477  
DB 418 olvpdxyLATVLMWKGKGTAFSAFVYVYTAELPTVVRNMGVGSSTASRLGSLIS 477  
QY 421 GLVPEDLHYISCTLVWVGKFGITLSAYSMVYVYTAELPTVVRNMGVGSSTASRLGSLIS 480  
DB 421 glvpEDLHYISCTLVWVGKFGITLSAYSMVYVYTAELPTVVRNMGVGSSTASRLGSLIS 480  
QY 478 PFYVVLGADRFPLPYILMGSLTILTAITLFLPESGTPLPDITDQMLRVKGMKHKRTPS 537  
DB 478 pfyvVLGADRFPLPYILMGSLTILTAITLFLPESGTPLPDITDQMLRVKGMKHKRTPS 537  
DB 481 PFYVYLGYAYDRPLPYILMGSLTILTAITLFLPESGTPLPDITDQMLRVKGMKHKRTPS 540  
QY 481 vylGADRFPLPYILMGSLTILTAITLFLPESGTPLPDITDQMLRVKGMKHKRTPS 540  
DB 479 vylGADRFPLPYILMGSLTILTAITLFLPESGTPLPDITDQMLRVKGMKHKRTPS 540  
QY 541 MLKGOERPTIKSTAF 557  
DB 536 dsmeetenpkl-ltaf 551



```

RESULT 9
AAV01651
ID AAV01651 standard; Protein: 553 AA.
XX
AC AAV01651;
XX
DT 23-JUN-1999 (first entry)
XX
DE A protein with cation transporting activity.
XX
KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
KM heart disease; cancer; anti-tumour drug; anticancer drug.
XX
OS Mus musculus.
XX
PN MO9913072-A1.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98MO-JP04009.
XX
PR 20-MAY-1998; 98JP-0156660.
XX
PR 08-SEP-1997; 97JP-0260972.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Nezu J, Oku A;
XX
DR WPI: 1999-215062/18.
DR N-PSDB; AAX2898.
XX
PT Genes homologous with organic cation transporters OCT1 and OCT2,
PT useful in design of new drugs for treatment of diseases due to
XX abnormality of the transporter functions
XX
PS Claim 1; Page 63-67; 97pp; Japanese.
XX
CC The present sequence represents a protein with cation transporting
CC activity. The genes are significantly homologous with organic cation
CC transporters OCT1 and OCT2. The genes may used in drug development,
CC particularly in the treatment of diseases due to abnormality of the
CC organic cation transporter functions e.g. fatty liver, heart diseases
CC and cancers, by controlling such as by inhibition or activation.
CC Administration of anti-tumour and anticancer drugs in combination with
CC a transporter protein inhibiting agent allows the agents to penetrate
CC into the diseased cells to enhance the drug action.
XX
SQ Sequence 553 AA;

Query Match 74.5%; Score 2148; DB 20; Length 553;
Best Local Similarity 72.2%; Pred. No. 2,5e-224;
Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;

```

```

DB 241 igymvlpjlfayfirdwrmlllaillpjlfcyplwffipesprwlisqrfaeaegliqka 300
QY 301 AKANGIWPSTIRDPSELQDLSKKOOSHNIIDLRTNIRWVTISIMLWMTISVGRG 360
DB 301 akmsivapagildpldelqelsikqkvilidlftrlnatltvmaawnlmltsyfa 360
QY 361 LSIDTPNLHGDIFFNCFLSAMVEVPAYVLAMLLQYLPRRYSNATPFGGSVLEFMQV 420
DB 361 lslnvplngdylnclisgllievpayfawlllrclpryliaaylfgwgvvllliqv 420
QY 421 PPDLYIATVLYVMWGEVTAFFSMVYVYTAELYPVFNMGVGSSTASRGSILSPYF 480
DB 421 pedynfvslgvlmglkfjgtsatfsmlyvfteelyplvrmnavgltmsarvgsilapyf 480
QY 481 VYLGAYDRFLPILMGSILTLFLFLPESGCTPLPDTIDMLRVKAMKRRKIPSTR 540
DB 481 vylgaynrllpyllmgslevlglitlflfpefgyvlpenlegmqkvirfgsk--xst 537
QY 541 MKDGOERPTILKSTAF 557
DB 538 vsvdreespkvl-itaif 553

RESULT 10
AAB20579
ID AAB20579 standard; Protein: 553 AA.
XX
AC AAB20579;
XX
DT 11-DEC-2000 (first entry)
XX
DE Mouse OCTN1 amino acid sequence.
XX
KW Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;
KW identification; regulator; carnitine transport.
XX
OS Mus musculus.
XX
PN MO200046368-A1.
XX
PD 10-AUG-2000.
XX
PF 04-FEB-2000; 2000MO-JP00619.
XX
PR 05-FEB-1999; 99JP-0028406.
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Nezu J, Ose A;
XX
DR WPI: 2000-586982/55.
XX
PT Organic cation transporter gene OCTN3 expressed in testis for
PT identification of regulators of carnitine transport for use as drugs
XX
PS Example 2; Fig 2; 58pp; Japanese.
XX
CC The present invention describes a mouse organic cation transporter
CC protein (OCTN3). Also described are: (1) a method for screening
CC compounds for their ability to regulate the transport of an organic
CC cation into the cell, by generating a cell expressing OCTN3 at the
CC cell membrane, contacting with the compound and organic cation, and
CC observing the degree of transport of the organic cation; and (2) a
CC method for screening compounds for their ability to be transported into
CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell
CC membrane, contacting with the compound and observing the degree of
CC transport of the compound. OCTN3 can be used for the identification of
CC regulators of the transport of organic cations (especially carnitine)
CC into cells by OCTN3, for use as drugs. The present sequence represents
CC the mouse OCTN1 amino acid sequence, which is used in an example from
CC the present invention.
XX
SQ Sequence 553 AA;

```

Query Match 74.5%; Score 2148; DB 21; Length 553;  
Best Local Similarity 72.2%; Pred. No. 2.5e-224;  
Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;

Oy 1 MRDYEVAAPFGEEMPFORLIFELLASIIINGFTGLSSVFLIMPEHRCRPPDANLSS 60  
Db 1 mrdyevafifgeempfifllifllsasilpnqfngmsvfliaqpehrclvpltnlss 60  
Oy 61 AMRNHTVRLRLDGRVHSCRRYRLATIANFSALGLEPGHDVDLGOLEQSCLDGWEFS 120  
Db 61 swrnhtvrlrlldgrvhsrryrlatiansfalsamglepgdvdldqlegscldgweyd 120  
Oy 121 ODVYSTVTETMNLVCEDDMKAPLISLFFYGVLLGSISQSLSDRGKRVLVVTMGQ 180  
Db 121 kdlfstltetmnlvceddmkpltslftvylcgstsvsglsdrgrkkrvlatmavq 180  
Oy 181 TGFSFLQFSKNFEMFVFLVVGNGQISNYVAFVLGTETILGKSVRIEFTLGVCIFYA 240  
Db 181 tgfstfvqlfscnemtvlfaivmgqisnyvafllgtelstksvrlfslgtvcfffa 240  
Oy 241 FGYNVLPFAVFIRDMRLMLVALTMPGVLCVAMWFIPESPRMLISQGREBEAVIIRKA 300  
Db 241 lqymvlpfayfirdmrlmlvaltmpgvlcvalmwfiepsprmlisqgrfaeeqllqka 300  
Oy 301 AKANGIVPSTIFDPESELODSSKKOOSHNLIDLRTMINIMVIMSIMLMTISVGYFG 360  
Db 301 akmsivapagldpdlqelnslkqkvllldlfrtnalctfvmavmwmllssvgyfa 360  
Oy 361 LSLDPNLGDFVNCFLSAMVEPAYVLAAMLQYLRPRXSMAATFLGSGVLLFFMQLV 420  
Db 361 lslvprnlhgdylnclfsisglievpayftawllrltpryiliegvlfgwggvllllqyv 420  
Oy 421 PPDLYLATVLVMGKFEVTAFAFSKVVYTAELFPTVVRNMGVGSSTASRLGSILSPYE 480  
Db 421 pedyfaylglvmlgkfgltsafmlyvftaelypclvrmavgltsmasvsgsilapyf 480  
Oy 481 VILGAYDFLPYILMGSLTITLITLTLFLPESFGPRLPDTIDOMLRVKGMRHRTPSHTR 540  
Db 481 vylgaydfnllpyllmgslvtlglitlflfespfgvcltpehlegmqkvrgitcgr--kst 537  
Oy 541 MLKDGQERPTILKSTAF 557  
Db 538 vsdreespkv1-itaef 553

RESULT 11  
ABB62279  
ID ABB62279 standard; Protein: 548 AA.  
XX ABB62279;  
DT 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 13629.  
XX Drosophila: developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX Drosophila melanogaster.  
OS  
PN WO200171042-A2.  
PD 27-SEP-2001.  
PF 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX

PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI, 2001-656860/75.  
XX N-PSDB; ABL06382.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 13629; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL1840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 548 AA;

Query Match 30.7%; Score 885; DB 22; Length 548;  
Best Local Similarity 35.1%; Pred. No. 5.7e-87;  
Matches 198; Conservative 106; Mismatches 206; Indels 54; Gaps 9;

Oy 4 YDEVATFLGEMGPFORLIFELLASIIIPNGFTGLSSVFLIAPENHRCRPP---DANLSS 59  
Db 4 ydevatflgfgpygkryyllclpalvacfbhlagvflakpfdccalpyengslyels 59  
Oy 60 SAMRNHTVRLRLDGRVHSCRRYRLATIANFSALGLEPGHDVDL-----GQL----- 108  
Db 63 phlnvlsyener-----csy-----dvdyteeylmgslpsrsn 97  
Oy 109 EQSSCLDGEFSDVDYLTIVTEMNLCEDDMKAPLISLFFYGVLLGSFTISQSLSDRG 168  
Db 98 ektlc-ssyyddskylnsavetmnlvcsrlsatsdelfmgvllgslfsgmsdkig 156  
Oy 169 RKNVLFVTMGQTFSEFLQFSKNFEMFVFLVVGNGQISNYVAFVLGTETILGKSVRI 228  
Db 157 rkptffaslvqlifgvlaavapeysyltsrmlvgatsgvlvaylalemgvsyr- 215  
Oy 229 IFSTLGVCIFYARGYWLPLFAFIRDMRLMLVALTMPGVLCVAMWFIPESPRMLISQ 288  
Db 216 lfagvamqmtfsvglmtagfayfndwrlqatclpgllficywllpessarvllmk9 275  
Oy 289 RFEBAEYIIRKAKANGIVPSTIFD-----PSELODSSKKOOSHNLIDLRTWNR 341  
Db 276 rkdeaivilekaakenkvevneieqlvdeavekkkgdemaasgaqavfdlllyrnlr 335  
Oy 342 MVTIMSLMTISVGYFLSLDTPNLGDFVNCFLSAMVEPAYVLAAMLQYLRPRY 401  
Db 336 rkclllffdwfnsgyuglswntlnlgngvlnfnmgsaveipgyltllfltnvgrts 395  
Oy 402 SMATLFLGSGVLLFMQDVPDLYLATVLVMGKRGCVTAFAFSKVVYTAELFPTVVRNM 461  
Db 396 lfsgtmnvaglsllatlfpsdmwllvacamgklatssygllylfaeagfprvtniv 455  
Oy 462 GVGVSSTASRLGSILSPFYUULGAYORFLPYILMGSLTITLITLTLFLPESFGPRLPDT 521  
Db 456 glgassmavarglilapylklldgeiwrpirlilcgalslvaagllslpeltinkmpet 515  
Oy 522 DOMLRVKGMRHRTPSHTRMLKDG 545  
Db 516 ed-----genfkkpqrqetaeeg 534

RESULT 12  
ABB68354

ID ABB68354 standard: Protein; 567 AA.  
 XX  
 AC ABB68354;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 31854.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001MO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EM;  
 DR N-PSDB; ABL12457.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 31854; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pot\_sequences.  
 XX

```

QY 293 AEIIRKAKNGIVP-----STFEDESE-----LQDSSKKQOQHNIIDLRTWN 339
Db 280 aiaamqkaarkinkveisdealseilldegeensekakqkledgeldeqpppswdlfcypn 339
QY 340 IRMYTINSIMLMWMTISVGEFGLSDTPENLBGDIFVNCFLSAMVEVPAYVAILMLLOYLPR 399
Db 340 lrrklllffldwltvltsgvyglswtanslgnvllnfvisgaveipayfflllltlhrwgr 399
QY 400 RYSNATALFLGGSVLLFMQLVPPDLTYLAVLVMWCKFGYTAAFSMWVYVYTAELPYTYR 459
Db 400 rslgcglvmaaglsllatvllpqmhbcllvacamlglklatasgyatvlytsaeqfcltvr 459
QY 460 NMGVSSTASRLSGLSIISPYFVYLGAADREPLPYTLMOSTLILAILTLFLESFGPLPD 519
Db 460 nvalgaasamarisgmapfnfiatiwkrpdlpdlcgslvagllslllpethnxpmle 519
QY 520 TLDDMLRYKSKMKHKRKTFSHRLMKDGE 547
Db 520 tl-----adgerfgkktkadvyletqge 542

```

RESULT 13  
ID AAM78367 standard; Protein; 577 AA.  
XX  
AC AAM78367;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human protein SEQ ID NO 1029.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN W0200157190-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 05-FEB-2001; 2001WO-US04098.  
XX  
PR 03-FEB-2000; 2000US-06A96914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhuo QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
DR WPI: 2001-476283/51.  
DR N-PSDB: AAK51500.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
PT  
XX  
XX Claim 20: Page 3266-3267; 6221pp; English.  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78373-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or



QY	118	EFSDQDYVSTIYTEENKIVCEDQKAPLTLISLFPVYGLGSLFSISGOLSPRFRKXVLEVTM	177
Db	113	iyqntkwtstactvqnnlvcotkklamlipqltmfgyllgsvtcfyfsdigrvynlwats	172
QY	178	GMQGFSEFLQIFSKNEMEFVYLVVGMGQISNVAAFPVLAGTEILGKSVRIIFSTLGVC	237
Db	173	ssmfllglaaafavdyvlfemaarffllamvasylyvgfyvmeflqmskr-twasvnlhs	231
QY	238	EYAFGMYVPLFPAAYTRIRW---RMLLYALMPGVLCVVALMWFIPSPSPRIISQGRFEAE	294
Db	232	ffvaygllylvalqylylvtwlygmllstvcvplficc---wlypevplwllsegryeaq	288
QY	295	VIIIRAKAKANGIVBSTIFDEPSLODI-----SSKKQOSHNIIDLRTWNIRMYTM	346
Db	289	klydlnakwn-----taasscklsellslslidgpyvsnpsrtevgqhnlsylfymwsllkrlt	344
QY	347	SIMLTMTIVGVEGSESLTPTPLMGHDIVNCFELNMEVEVAYVLAWLMLLYOYPRRYSNATA	406
Db	345	vwlwtgtslgytslslsvnlsgnqeylnllfgvveipayctfvciamdkygrttvlay	404
QY	407	LELGSGVLLFEMOLVPELDLYLATVLVVNGKFGVTAAFSMVYVYVYAEVPTVVRNMGVYS	466
Db	405	lfscalacgvnmvlpqkhyllgvvtamvqkfaigaafgillylvtelypclivrslawsg	464
QY	467	STASRLGSLTSPYFYVLYGAYRFLPYTLMGSLTITLITLFLPSPFGPPLDPTIDMLR	526
Db	465	smwcrllasllapfsvdsislwflfpqlfygtmalisgvtllklpeltgkrlatlwaeak	524
QY	527	VKGMKHKRT 535	
Db	525	lesenesks 533	
RESULT	15		
ID	AAAM00930		
AA	AAAM00930 standard; Protein: 584 AA.		
XX	AAAM00930;		
XX	01-OCT-2001 (first entry)		
XX			
XX	Human bone marrow protein, SEQ ID NO: 406.		
XX			
XX	Human: bone marrow; antiinflammatory; cyostatic; neuroprotective;		
XX	antiviral; antibacterial; antitungal; anti-HIV; hemostatic;		
XX	immunosuppressive; gene therapy; cytokine cell proliferation;		
XX	cell differentiation modulator; immune disorder; infection; cancer;		
XX	human immunodeficiency virus; HIV; autoimmune disorder; haemophilija.		
OS	Homo sapiens.		
XX			
XX	MO200153453-AA2.		
XX			
XX	26-JUL-2001.		
XX			
XX	23-DEC-2000; 2000WO-US34960.		
XX			
XX	21-JAN-2000; 2000US-0488725.		
XX	25-APR-2000; 2000US-0552317.		
XX	09-JUL-2000; 2000US-0598042.		
XX	19-JUL-2000; 2000US-0620312.		
XX	03-AUG-2000; 2000US-0653450.		
XX	14-SEP-2000; 2000US-0662191.		
XX	19-OCT-2000; 2000US-0693036.		
XX	30-NOV-2000; 2000US-0250583.		
XX			
XX	(HYSE-) HYSEQ INC.		
XX			
XX	Ford JB, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;		
XX	Ren F, Wang J, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J;		
XX	Zhao QA, Zhou P, Drmanac RT.		
XX	WPI; 2001-488707/53.		

DR N-PSDB; AAH90049.

PT Novel bone-marrow-expressed polynucleotides and polypeptides, useful  
PT for treating e.g. cancer and immune deficiency disorders -

XX  
XX  
PS Claim 10; Page 490-491; 648pp; English.

XX The present sequence is one of 251 novel human polypeptides encoded  
CC by a bone marrow-expressed polynucleotide. The polynucleotide and the  
CC polypeptide encoded by it are useful in the treatment of various  
CC immune deficiencies and disorders. The deficiencies and disorders may  
CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal  
CC infection (e.g. haemophilias), inhibition of tumour cell proliferation,  
CC suppression of an inflammatory response or treatment of a nervous  
CC system disorder such as Alzheimer's disease. Detection of the presence  
CC or increased expression of the polynucleotide or the protein it  
CC encodes is useful for the diagnosis and/or prognosis of one  
CC or more types of cancer. The polynucleotide and polypeptide can be  
CC used as nutritional sources or supplements and in the screening of  
CC chemical compounds as potential drugs.

XX  
SQ Sequence 584 AA:

Query Match 27.2%; Score 784; DB 22; Length 584;  
Best Local Similarity 33.1%; Pred. No. 6e-76;  
Matches 187; Conservative 110; Mismatches 220; Indels 48; Gaps 10;

OY 2 RDYDEVATFAGLGEWGPORLIFFLLSASIIIPNGTGLSSVEPLATPEHRCRP----- 53  
| : : : | : | : | : : : | : | : | : | : | : | : | : | :  
Dd rhfeqiydhvghygrfygrylylficaftqncjibhylaefmgvyphrcrpnqvavf 69  
OY 54 -----DAANLSAMENHNHYPRLRDGR--EYPHSRCRYRLATIANFSALGLEPGND 102  
| : : : | : | : : | : : : | : | : | : | : | : | : | : | :  
Dd hnhsmwledtgallssgdqdyvvqlngelweistocrnk-----enstsiye---- 121  
OY 103 VDLQQLBOESCLDGWEFSODVYLSTIVTEWMNLCEDDMKAPLTISLFYGVLLGSFISSQ 162  
| : : | : | : : | : : | : | : | : | : | : | : | : | : | :  
Dd 122 -ytgskkfefpcvdgyldqntwkstcavtqwlnocdrkwlamlqplmfvgvlisvtfgy 180  
OY 163 LSDRRGRNVLFTMGMGNGEFLIOFSKNEMFPYLVLYVMGQISNVYAVALGTETL 222  
| | | | : | : | : | : : | : : | : | : | : | : | : | :  
Dd ftsdlitrrivvaawssmflgliaafadvdyttlmaarfiamasylyvgfvymeti 240  
OY 223 GKSVRIIFSTLGVCIFYAFGYWVLPFAFYRDM--RMILVALTMPGVLCVALMWFIE 279  
| : : : : | : : | : | : | : | : | : | : | : | : | : | :  
Dd 241 gmkser-twaashhsffavgtllvaltylvrtwmlygmilstvtfpfllcc---wvpe 296  
OY 280 SPRMLISOGPREAEAYIRKAANKNIWSPRIIDPSELDL-----SSKKOOSHNI 331  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Dd 297 tpfwllseegryeeagklvdimakn-----raasccklselislidlgpvsnspvevqhnl 352  
OY 332 LDLERTNRKWTMTSIMLMNTISVGEGSLIDTPRELDIGDIYNVCFLSNAVEVPVAVIAM 391  
| : | : | : | : : | : | : | : | : | : | : | : | : | : | : | :  
Dd 353 sylfnyswikrlwtwllwfctsglgyfsfnsnvlnagnneylnllfiylgvelpaytlvc 412  
OY 392 LLIDYLRPRSMATAFLFGSGVYLEFMQADPOLPYLAVALVIAWVK--RGVYAAASMYIYT 450  
| : : | : : | : | : | : | : | : : | : | : | : | : | : | : | :  
Dd 413 iamkvvgrtevlaysifcsalgaeyvmvtpdqhnylllygtvammygklipiaaaqillylt 472  
OY 451 AELPVTVARNMGVGSVASLRGSIIISPYEVYLGAVDRFLPYIIMGSLTLTAITLFLP 510  
| | | | : | : | : | : | : | : | : | : | : | : | : | : | :  
Dd 473 aelyplvtirsiaavgswmcorlasiafpdsvlssiwiifpqflvsgimajslsvcltkip 532  
OY 511 ESFGPLPDTRIDOKLRVKGKMHKKT 535  
| : | : | : | : : | : : | : | : | : | : | : | : | : | :  
Dd 533 etligrattatweeaakieseneaks 557



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 02:03:56 ; Search time 37.9 Seconds  
(Without alignments)  
358.972 Million cell updates/sec

Title: US-09-521-195B-3  
Perfect score: 2883  
Sequence: 1 MRDYDEVTAFLGEMGFQRL.....HTRMLKDGGERPTLIKSTAF 557

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	726	25.2	555	US-08-501-572-3	Sequence 3, Appli
2	726	25.2	555	US-09-040-444-3	Sequence 3, Appli
3	718	24.9	556	US-08-501-572-1	Sequence 1, Appli
4	718	24.9	556	US-09-040-444-1	Sequence 1, Appli
5	714.5	24.8	553	US-08-501-572-2	Sequence 2, Appli
6	714.5	24.8	553	US-09-040-444-2	Sequence 2, Appli
7	652.5	22.6	537	US-08-647-397-2	Sequence 2, Appli
8	302.5	10.5	520	US-08-964-127-2	Sequence 2, Appli
9	302.5	10.5	520	US-09-496-692-2	Sequence 2, Appli
10	234	8.1	494	US-09-031-392-5	Sequence 5, Appli
11	234	8.1	494	US-09-299-549-5	Sequence 5, Appli
12	234	8.1	494	US-09-610-417-5	Sequence 5, Appli
13	214.5	7.4	286	US-08-964-127-4	Sequence 4, Appli
14	214.5	7.4	286	US-09-496-692-4	Sequence 4, Appli
15	208	7.2	492	US-08-355-844-3	Sequence 3, Appli
16	208	7.2	492	PCT-US95-16126-3	Sequence 3, Appli
17	203	7.0	493	US-09-031-392-10	Sequence 10, Appli
18	203	7.0	493	US-09-299-549-10	Sequence 10, Appli
19	203	7.0	493	US-09-610-417-10	Sequence 10, Appli
20	198	6.9	109	US-08-647-397-4	Sequence 4, Appli
21	193	6.7	509	US-09-031-392-6	Sequence 6, Appli
22	193	6.7	509	US-09-299-549-6	Sequence 6, Appli
23	193	6.7	509	US-09-610-417-6	Sequence 6, Appli
24	186	6.5	500	US-09-031-392-7	Sequence 7, Appli
25	186	6.5	500	US-09-299-549-7	Sequence 7, Appli
26	186	6.5	500	US-09-610-417-7	Sequence 7, Appli
27	184	6.4	563	US-09-031-392-2	Sequence 2, Appli

28	184	6.4	563	4	US-09-299-549-2	Sequence 2, Appli
29	184	6.4	563	4	US-09-610-417-2	Sequence 2, Appli
30	175	6.1	488	2	US-08-928-692-11	Sequence 11, Appli
31	171	5.9	322	4	US-08-964-127-6	Sequence 6, Appli
32	171	5.9	322	4	US-09-496-692-6	Sequence 6, Appli
33	168.5	5.8	534	2	US-09-031-392-4	Sequence 4, Appli
34	168.5	5.8	534	4	US-09-299-549-4	Sequence 4, Appli
35	168.5	5.8	534	4	US-09-610-417-4	Sequence 4, Appli
36	155	5.4	383	2	US-09-031-392-3	Sequence 3, Appli
37	155	5.4	383	4	US-09-299-549-3	Sequence 3, Appli
38	155	5.4	383	4	US-09-610-417-3	Sequence 3, Appli
39	153.5	5.3	524	2	US-08-928-692-12	Sequence 12, Appli
40	149	5.2	584	2	US-08-928-692-13	Sequence 13, Appli
41	131	4.5	488	2	US-08-928-692-10	Sequence 10, Appli
42	110	3.8	473	1	US-08-597-236-13	Sequence 13, Appli
43	110	3.8	473	1	US-08-746-682A-13	Sequence 13, Appli
44	104.5	3.6	1835	5	US-08-836-325-15	Sequence 15, Appli
45	101.5	3.5	341	1	US-08-423-564-5	Sequence 5, Appli

## ALIGNMENTS

RESULT 1  
US-08-501-572-3  
; Sequence 3, Application US/08501572  
; Patent No. 6063623  
GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; TITLE OF INVENTION: Transport protein which effects the  
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals,  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/501,572  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toohey, Kimberlin M  
; REGISTRATION NUMBER: 35,391  
; REFERENCE/DOCKET NUMBER: 02481.1453-00000  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 555 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-501-572-3

Query Match 25.2%; Score 726; DB 3; Length 555;  
Best Local Similarity 35.0%; Pred. No. 5.9e-68;  
Matches 187; Conservative 98; Mismatches 223; Indels 26; Gaps 14;  
QY 5 DEVTAFLGEMGFQRLIF---LLISATIPNGTGLSSVFLIATPEHRCVPDAANLS-- 59





```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/501,572
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Toohey, Kimberlin M
REGISTRATION NUMBER: 35,391
REFERENCE/DOCKET NUMBER: 02481.1453-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-501-572-1

```

```

Query Match          24.9%: Score 718; DB 3; Length 556;
Best Local Similarity 34.0%: Pred. No. 4.1e-67;
Matches 194; Conservative 96; Mismatches 229; Indels 52; Gaps 14;

```

```

QY 1 MRDYEYTAFLGEMGPORLIFELL---SASIIPIGFTGLSSVFLIATPEHRCRPDPAAN 57
   1 MPTVDVLEQVGEFGMGKQAFLLCLISASLAPI-YVGI--VFLGFTPHGYQNGPVAAE 57
Db 1 MPTVDVLEQVGEFGMGKQAFLLCLISASLAPI-YVGI--VFLGFTPHGYQNGPVAAE 57
QY 58 LSS--AMR-----NHTVP-LRLDGRVPHSCRRRL-----ATIANFSAL 95
   58 LSS--AMR-----NHTVP-LRLDGRVPHSCRRRL-----ATIANFSAL 95
Db 58 LSS--AMR-----NHTVP-LRLDGRVPHSCRRRL-----ATIANFSAL 95
QY 96 GLEPGRDVDLGLQLEQSCLDGMEFSQDYLSTIVTEEMNLVCEDDMKAPLTISLFFVGLL 155
   96 GLEPGRDVDLGLQLEQSCLDGMEFSQDYLSTIVTEEMNLVCEDDMKAPLTISLFFVGLL 155
Db 96 GLEPGRDVDLGLQLEQSCLDGMEFSQDYLSTIVTEEMNLVCEDDMKAPLTISLFFVGLL 155
QY 118 PLGP-----CEHGMVY--DTPGSSIVTEFNLVCGDAKKVDLFQSCVNLGFFL 162
   118 PLGP-----CEHGMVY--DTPGSSIVTEFNLVCGDAKKVDLFQSCVNLGFFL 162
Db 118 PLGP-----CEHGMVY--DTPGSSIVTEFNLVCGDAKKVDLFQSCVNLGFFL 162
QY 156 GSFISGQLSDRGRKKNVLFYTMGMQTFSEFLQIFSKNEFVVLFLVYMGQISNYAAAF 215
   156 GSFISGQLSDRGRKKNVLFYTMGMQTFSEFLQIFSKNEFVVLFLVYMGQISNYAAAF 215
Db 156 GSFISGQLSDRGRKKNVLFYTMGMQTFSEFLQIFSKNEFVVLFLVYMGQISNYAAAF 215
QY 163 GSLVGYIADRGGRKLCILVTLVTSVSGVLTAVAPDYSMLFRLQGWGWSGWSVSGY 222
   163 GSLVGYIADRGGRKLCILVTLVTSVSGVLTAVAPDYSMLFRLQGWGWSGWSVSGY 222
Db 163 GSLVGYIADRGGRKLCILVTLVTSVSGVLTAVAPDYSMLFRLQGWGWSGWSVSGY 222
QY 216 VLGTETLLKSVRIITSTGLVCIFYAF--GYMYLPLFAFYFIDRMRLMLVALTMPGVLYAL 273
   216 VLGTETLLKSVRIITSTGLVCIFYAF--GYMYLPLFAFYFIDRMRLMLVALTMPGVLYAL 273
Db 216 VLGTETLLKSVRIITSTGLVCIFYAF--GYMYLPLFAFYFIDRMRLMLVALTMPGVLYAL 273
QY 223 TLTFEFGSGYR--RTAIIYQMAFTVGLVGLAGVAAIAPDMRWQLAVSLPTFFELLY 279
   223 TLTFEFGSGYR--RTAIIYQMAFTVGLVGLAGVAAIAPDMRWQLAVSLPTFFELLY 279
Db 223 TLTFEFGSGYR--RTAIIYQMAFTVGLVGLAGVAAIAPDMRWQLAVSLPTFFELLY 279
QY 274 WMFIPESPRLWISQGRFEAEVLIIRKAKANGIVPSTIPDSELDLSKKQOSHNTLD 333
   274 WMFIPESPRLWISQGRFEAEVLIIRKAKANGIVPSTIPDSELDLSKKQOSHNTLD 333
Db 274 WMFIPESPRLWISQGRFEAEVLIIRKAKANGIVPSTIPDSELDLSKKQOSHNTLD 333
QY 280 YMFVPESEPRMLLSQKRTTRAVRIMEQIAQKNGKVPADLMKLCLEEDASEKR--SPSFAD 337
   280 YMFVPESEPRMLLSQKRTTRAVRIMEQIAQKNGKVPADLMKLCLEEDASEKR--SPSFAD 337
Db 280 YMFVPESEPRMLLSQKRTTRAVRIMEQIAQKNGKVPADLMKLCLEEDASEKR--SPSFAD 337
QY 334 LRTWNIRAVTITMSIMLMTISVGYFGLSDTPNLHGDIYFNCFLSAMVEVPAYVLAWL 393
   334 LRTWNIRAVTITMSIMLMTISVGYFGLSDTPNLHGDIYFNCFLSAMVEVPAYVLAWL 393
Db 334 LRTWNIRAVTITMSIMLMTISVGYFGLSDTPNLHGDIYFNCFLSAMVEVPAYVLAWL 393
QY 338 LFRTPNLRKHTVILMYLWFSCAVLQGLIMHVAGATGANLYLDFEYSSLVPEPAFFIIVT 397
   338 LFRTPNLRKHTVILMYLWFSCAVLQGLIMHVAGATGANLYLDFEYSSLVPEPAFFIIVT 397
Db 338 LFRTPNLRKHTVILMYLWFSCAVLQGLIMHVAGATGANLYLDFEYSSLVPEPAFFIIVT 397
QY 394 LQYLPRYSMAATALFLGGSVLLFMOLVPPDLYLATVLMVNGFGVTAASMYVYVTAEL 453
   394 LQYLPRYSMAATALFLGGSVLLFMOLVPPDLYLATVLMVNGFGVTAASMYVYVTAEL 453
Db 394 LQYLPRYSMAATALFLGGSVLLFMOLVPPDLYLATVLMVNGFGVTAASMYVYVTAEL 453
QY 398 IDRIGRIYPIAASNLVYGAACLMIPIPHLHMLNVTIACLRMGATVILEMVCINVAEL 457
   398 IDRIGRIYPIAASNLVYGAACLMIPIPHLHMLNVTIACLRMGATVILEMVCINVAEL 457
Db 398 IDRIGRIYPIAASNLVYGAACLMIPIPHLHMLNVTIACLRMGATVILEMVCINVAEL 457
QY 454 YPTVVRNMGVGSSTASRLSIIISPFYVY--LGAYDRFLPYITLMSFTIITAILFTFLPS 512
   454 YPTVVRNMGVGSSTASRLSIIISPFYVY--LGAYDRFLPYITLMSFTIITAILFTFLPS 512
Db 454 YPTVVRNMGVGSSTASRLSIIISPFYVY--LGAYDRFLPYITLMSFTIITAILFTFLPS 512
QY 458 YPFIRNLGMMVCSALCDLQIGITPPFVFRIMLMEWQALPILILGVYGLTAGAATLILPEL 517
   458 YPFIRNLGMMVCSALCDLQIGITPPFVFRIMLMEWQALPILILGVYGLTAGAATLILPEL 517
Db 458 YPFIRNLGMMVCSALCDLQIGITPPFVFRIMLMEWQALPILILGVYGLTAGAATLILPEL 517
QY 513 FGPPPLDPTIDMLRVKGMKHKRTPSHTRMK 543
   513 FGPPPLDPTIDMLRVKGMKHKRTPSHTRMK 543
Db 513 FGPPPLDPTIDMLRVKGMKHKRTPSHTRMK 543
QY 518 KVALPETIEAEMLT-GRRSKAKENTIIYLQ 547
   518 KVALPETIEAEMLT-GRRSKAKENTIIYLQ 547
Db 518 KVALPETIEAEMLT-GRRSKAKENTIIYLQ 547

```

```

RESULT 4
US-09-040-444-1
Sequence 1, Application US/09040444
Patent No. 6063766
GENERAL INFORMATION:
APPLICANT: Koepsell, Hermann
APPLICANT: Grundeman, Dirk

```

```

APPLICANT: Gorboulev, Valentin
TITLE OF INVENTION: Transport of protein which effects the
TITLE OF INVENTION: Transport of cationic xenobiotics and/or Pharmaceuticals,
TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Finegan, Henderson, Farabow, Garrett & Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,444
FILING DATE: March 18, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Steven P
REGISTRATION NUMBER: 41,225
REFERENCE/DOCKET NUMBER: 2481.1453-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4000
TELEFAX: (202)408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-040-444-1

```

```

Query Match          24.9%: Score 718; DB 3; Length 556;
Best Local Similarity 34.0%: Pred. No. 4.1e-67;
Matches 194; Conservative 96; Mismatches 229; Indels 52; Gaps 14;

```

```

QY 1 MRDYEYTAFLGEMGPORLIFELL---SASIIPIGFTGLSSVFLIATPEHRCRPDPAAN 57
   1 MPTVDVLEQVGEFGMGKQAFLLCLISASLAPI-YVGI--VFLGFTPHGYQNGPVAAE 57
Db 1 MPTVDVLEQVGEFGMGKQAFLLCLISASLAPI-YVGI--VFLGFTPHGYQNGPVAAE 57
QY 58 LSS--AMR-----NHTVP-LRLDGRVPHSCRRRL-----ATIANFSAL 95
   58 LSS--AMR-----NHTVP-LRLDGRVPHSCRRRL-----ATIANFSAL 95
Db 58 LSS--AMR-----NHTVP-LRLDGRVPHSCRRRL-----ATIANFSAL 95
QY 96 GLEPGRDVDLGLQLEQSCLDGMEFSQDYLSTIVTEEMNLVCEDDMKAPLTISLFFVGLL 155
   96 GLEPGRDVDLGLQLEQSCLDGMEFSQDYLSTIVTEEMNLVCEDDMKAPLTISLFFVGLL 155
Db 96 GLEPGRDVDLGLQLEQSCLDGMEFSQDYLSTIVTEEMNLVCEDDMKAPLTISLFFVGLL 155
QY 118 PLGP-----CEHGMVY--DTPGSSIVTEFNLVCGDAKKVDLFQSCVNLGFFL 162
   118 PLGP-----CEHGMVY--DTPGSSIVTEFNLVCGDAKKVDLFQSCVNLGFFL 162
Db 118 PLGP-----CEHGMVY--DTPGSSIVTEFNLVCGDAKKVDLFQSCVNLGFFL 162
QY 156 GSFISGQLSDRGRKKNVLFYTMGMQTFSEFLQIFSKNEFVVLFLVYMGQISNYAAAF 215
   156 GSFISGQLSDRGRKKNVLFYTMGMQTFSEFLQIFSKNEFVVLFLVYMGQISNYAAAF 215
Db 156 GSFISGQLSDRGRKKNVLFYTMGMQTFSEFLQIFSKNEFVVLFLVYMGQISNYAAAF 215
QY 163 GSLVGYIADRGGRKLCILVTLVTSVSGVLTAVAPDYSMLFRLQGWGWSGWSVSGY 222
   163 GSLVGYIADRGGRKLCILVTLVTSVSGVLTAVAPDYSMLFRLQGWGWSGWSVSGY 222
Db 163 GSLVGYIADRGGRKLCILVTLVTSVSGVLTAVAPDYSMLFRLQGWGWSGWSVSGY 222
QY 216 VLGTETLLKSVRIITSTGLVCIFYAF--GYMYLPLFAFYFIDRMRLMLVALTMPGVLYAL 273
   216 VLGTETLLKSVRIITSTGLVCIFYAF--GYMYLPLFAFYFIDRMRLMLVALTMPGVLYAL 273
Db 216 VLGTETLLKSVRIITSTGLVCIFYAF--GYMYLPLFAFYFIDRMRLMLVALTMPGVLYAL 273
QY 223 TLTFEFGSGYR--RTAIIYQMAFTVGLVGLAGVAAIAPDMRWQLAVSLPTFFELLY 279
   223 TLTFEFGSGYR--RTAIIYQMAFTVGLVGLAGVAAIAPDMRWQLAVSLPTFFELLY 279
Db 223 TLTFEFGSGYR--RTAIIYQMAFTVGLVGLAGVAAIAPDMRWQLAVSLPTFFELLY 279
QY 274 WMFIPESPRLWISQGRFEAEVLIIRKAKANGIVPSTIPDSELDLSKKQOSHNTLD 333
   274 WMFIPESPRLWISQGRFEAEVLIIRKAKANGIVPSTIPDSELDLSKKQOSHNTLD 333
Db 274 WMFIPESPRLWISQGRFEAEVLIIRKAKANGIVPSTIPDSELDLSKKQOSHNTLD 333
QY 280 YMFVPESEPRMLLSQKRTTRAVRIMEQIAQKNGKVPADLMKLCLEEDASEKR--SPSFAD 337
   280 YMFVPESEPRMLLSQKRTTRAVRIMEQIAQKNGKVPADLMKLCLEEDASEKR--SPSFAD 337
Db 280 YMFVPESEPRMLLSQKRTTRAVRIMEQIAQKNGKVPADLMKLCLEEDASEKR--SPSFAD 337
QY 334 LRTWNIRAVTITMSIMLMTISVGYFGLSDTPNLHGDIYFNCFLSAMVEVPAYVLAWL 393
   334 LRTWNIRAVTITMSIMLMTISVGYFGLSDTPNLHGDIYFNCFLSAMVEVPAYVLAWL 393
Db 334 LRTWNIRAVTITMSIMLMTISVGYFGLSDTPNLHGDIYFNCFLSAMVEVPAYVLAWL 393
QY 338 LFRTPNLRKHTVILMYLWFSCAVLQGLIMHVAGATGANLYLDFEYSSLVPEPAFFIIVT 397
   338 LFRTPNLRKHTVILMYLWFSCAVLQGLIMHVAGATGANLYLDFEYSSLVPEPAFFIIVT 397
Db 338 LFRTPNLRKHTVILMYLWFSCAVLQGLIMHVAGATGANLYLDFEYSSLVPEPAFFIIVT 397
QY 394 LQYLPRYSMAATALFLGGSVLLFMOLVPPDLYLATVLMVNGFGVTAASMYVYVTAEL 453
   394 LQYLPRYSMAATALFLGGSVLLFMOLVPPDLYLATVLMVNGFGVTAASMYVYVTAEL 453
Db 394 LQYLPRYSMAATALFLGGSVLLFMOLVPPDLYLATVLMVNGFGVTAASMYVYVTAEL 453
QY 398 IDRIGRIYPIAASNLVYGAACLMIPIPHLHMLNVTIACLRMGATVILEMVCINVAEL 457
   398 IDRIGRIYPIAASNLVYGAACLMIPIPHLHMLNVTIACLRMGATVILEMVCINVAEL 457
Db 398 IDRIGRIYPIAASNLVYGAACLMIPIPHLHMLNVTIACLRMGATVILEMVCINVAEL 457

```

Query Match	24.8%;	Score 714.5;	DB 3;	Length 55
-------------	--------	--------------	-------	-----------

Best Local Similarity 35.4%; Pred. No. 9.6e-67;  
Matches 194; Conservative 87; Mismatches 216; Indels 51; Gaps 13;

```
QY 1 MRDDEVATLGEWGPOR---LIFFLLSASIPNGFTGSSVFLATPEHRCRVDAAN 57
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1 MPYVDLLEQVSGSGNFQKAPLILCLLSAEP---ICVGIYLFQFTPHHCQSGVAE 57
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 58 LSS---AMR---NHTVPLRLDRGREVPHSCRRY-----LATIA-NFSALG 96
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 58 LSQRCSGSPAELNYTPVGLGPAGEAFLGCRREYEDVMNQASLSCVDPILASLATN 117
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 97 LEEGRVDVLGQLESDQWEEFSDVYISTYTEMNLVCEDEWKAPLITISFFVVLG 156
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 118 LGR-----CQDSWYV--DTPGSSIVTEFNLVCAWSKLDLQSCINAGFFFG 162
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 157 SFISGQSDREGRKNVLFVWGMQTFSGFLOIFSKNPEWVLFVVGQISNYAAVY 216
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 163 SLGVGFADFGKRLCLLGLVLANVSVGLMAFSPNMSKLERLLQGLVSKGNMAGYT 222
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 217 LGTELLGKSVRIIFSTLGVCIFFYAF--GYVWLPPLFAVFIIDMRMLVALMPGLCALM 274
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 223 LIEFVSGSSR---RTVAIMYQMAFVYGLVALGLAVLPHMRWLDLAVSLPTFLFLY 279
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 275 WFIPESEKRLISGREGREAVIIRKAAKANGIYVPESTIEDPSLQDLSKKQOSHILDL 334
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 280 WCPSESEKRLISGREGREAVIIRKAAKANGIYVPESTIEDPSLQDLSKKQOSHILDL 337
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 335 LRTWNIWMTVMSIMLMTITSVGYFGLSDTPRLHGDIFNCLSLAMVEVPAYVALMLL 394
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 338 FRIPLRUKRTFIIMLYMTFSDVLYOGLILMGATSGNLYIDLYLSALVELPGAFIALIT 397
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 395 QYLRKRYSMATALELGLSVLLFMQLVPPDLXYLATVLMVWKGFGVTAASMYVYAEY 454
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 398 DRYGRIVPAVSNMLAGAACTV-FISPDLMHNTIIIMCGRMGITAIQMICLVNAELY 456
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 455 PTYVNNMGVSVSTASRLGSLSPRYV-LGAYDRPLPTLMGSLITLAILTLPLPEP 513
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 457 PTYVNNMGVSVSTASRLGSLSPRYV-LGAYDRPLPTLMGSLITLAILTLPLPEP 516
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 514 GTPLEPDT 521
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 517 GDALPEM 524
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 7  
US-08-647-397-2  
; Sequence 2, Application US/08647397  
; Patent No. 5972702  
; GENERAL INFORMATION:  
; APPLICANT: Beier, David R.  
; APPLICANT: Beier, Kevin P.  
; TITLE OF INVENTION: OSTEOCLAST TRANSPORTER  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/647,397  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gates, Edward R.  
; REGISTRATION NUMBER: 31,616  
; REFERENCE/DOCKET NUMBER: B0801/7048

TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-720-3500  
; TELEFAX: 617-720-2441  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 537 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-647-397-2

Query Match 22.6%; Score 652.5; DB 2; Length 537;  
Best Local Similarity 29.9%; Pred. No. 3.2e-60;  
Matches 167; Conservative 104; Mismatches 246; Indels 41; Gaps 12;

```
QY 4 YDEVTAFLGEMGPORLIFFLLSASIPNGFTGSSVFLATPEHRCRVDAANISSAMR 63
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3 FSEIIDRVGSMGPFQYLHVTLALPILIGIANHNLQIIFTATTPDHHCRRPNASL-BFW 60
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 64 NHTVPLRLDRGREVPHSCRRY---RLATIANFSALGEPGRVDVLGQLESDQWEEF 120
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 61 --VPLG-PNGK--PEKCLRFVHLPMASLPDGTGATP-----CUDGWITN 102
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 121 QDVLSTYTEMNLVCEDEWKAPLITISFFVGLGFSISGQSDREGRKNVLFVWGMQ 180
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 103 ST--RDTIYTEMNDLVCGSNKLEKMAQSVFMAGILVGPVFGELSDRFGRKPLIWSYLL 160
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 181 TGFSEFLOIFSKNPEWVLFVVGQISNYAAVAFVIGTEILGKSVRIIFS-TLGVCTEY 239
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 161 MASGSSAAFPSPFLVYMIFFRFLGCGSIGISLSTIILNVEWVPSTRAISSITIGYC--Y 218
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 240 AFGVWLPPLFAVFIIDMRMLVALMPGLCALMFIPESEKRLISGREGREAVIIRK 299
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 219 TTGQFTLELAAYVQKMLQDLSVAAPFISLSMWPESTRMLVLSGKFSRAKTLTKR 278
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 300 AKANGIYVPESTIEDPSL-----ODLSKKQOSHILDLRTWNIWMTVMSIMLMTI 354
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 279 VATFNGKKEGEKLVLEELKFNLOKDITSAYK-YGLSDLFVRSILRVATFCLSLAWPAT 337
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 355 SVGYFGLSDTPRLHGDIFNCLSLAMVEVPAYVALMLLOYLRRKSMATALELGLSVL 414
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 338 GRAYVSLAMGVEEFGVNIYIIQIFGVDIPAKFTIISISYLGRRITQGLLILAGVAI 397
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 415 LFMQLVPPDLXYLATVLMVWKGFGVTAASMYVYTAELPYVRNMGVSVSTASRLGS 474
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 398 LILIVSSFMQLRTPALAVFGKGLSGSFSCLELYTSELPVLYKQGMGISTIMARVGS 457
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 475 ILSPEYVYLGAAYDRPLPYILMGLTITLAILTLPLPESEFGTLPDPTD-----QMLRV 527
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 458 MIAPLVKLTIGELQPIPIPVIFMTMLLGGSAFLELTLNRPLETIEDIDQWYQTKKT 517
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 528 KGMKHKRTPSHTRMLKDG 545
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 518 KOEPEAKASQTIPLKGTG 535
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 8  
US-08-964-127-2  
; Sequence 2, Application US/08964127  
; Patent No. 6277565  
; GENERAL INFORMATION:  
; APPLICANT: Grandearl, Andrew David John  
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804





[illegible]



```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 492 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   ORIGINAL SOURCE:
;   ORGANISM: Human
;   FEATURE:
;     NAME/KEY: Peptide
;     LOCATION: 1..492
;   OTHER INFORMATION: Facilitative glucose transporter
;   OTHER INFORMATION: Glut protein
US-08-355-844-3
```

Query Match 7.2%; Score 208; DB 2; Length 492;  
Best Local Similarity 22.2%; Pred. No. 2.3e-13;  
Matches 108; Conservative 87; Mismatches 165; Indels 126; Gaps 22;

```
QY 119 FSQDYLSTIVTEWNVCDDBKAPLITSLFFVGLSGISGQSDRRGRK-----NV 172
Db 52 YGESILPTLTLLWS-----LSVAIFSVGMISFSVGLFVNRGRNSMLMMNL 101
QY 173 LEVTMGMTGFSEFLQIFSKFEMFVVLFEVLGM--GOISNYVAFAVLGTEILGKSVRIIF 230
Db 102 LAFVSAVLMGFESKL---GKSFEMILGRIIGVYGLTGTGFPMYV--GEVSPFAFGAL 156
QY 231 STL-----GVCIFYAGCYMVLPLFAFYIRD-----WRMLVALTMPGVLCVALMW 275
Db 157 GRLHOLGIVVGILIAOVEGL-----DSINGNKDLMPLLSIIFIPALLQCIVIP 205
QY 276 FIPESPRW-LISQGFEEAEVLIIRAKANGIVPSTIFDSELDLSKKQOSH----- 329
Db 206 FCPESPRLILIRNENNRKASVKKLRGTADY-----HDLOEMKESROMRE 254
QY 330 ---NILDLRTWNIIMVTIMSIMMT-----ISVGFGLSLDTPNLHGDIFVNCPLSA 380
Db 255 KVTILELFRSPAYRQPLIAVAVQLSQQLSGINAVFYSTSI-----FEKA 301
QY 381 MVEVPAY-----VIAMLLQYLPRR--YSMATALFLGGSVLLFMQLY-----P 421
Db 302 GVOQEPVATIGSGIVNIAFTVVSLEFVERAGRRLHLGLAGMAGQALIMTIALALEQL 361
QY 422 PDLVYLAFLVWVGKFGVTAAS-----MYVYTAELYPJVVRNMGVCSSTASRLGS-I 475
Db 362 PMMSYLSIAI-----FGVAFFEVGPGPIPMFIVALESQGPRAIIVAGFSNMTSNFI 417
QY 476 LSPYFVYL---GAYDRFLPYILMGSLTILAILTLF-LPESFGTPLPDTIDOMLRVKG 530
Db 418 VMCFQYVEOLGPGY---VFIIFFVLVLEFIIRYFVKVPEMKGTDEIASGFRQGGAS 473
QY 531 KHRKTP 536
Db 474 QSDKTP 479
```

Search completed: July 17, 2002, 02:03:58  
Job time: 5774 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 02:05:09 ; Search time 58.69 Seconds  
(without alignments)  
911.940 Million cell updates/sec

Title: US-09-521-195B-3  
Perfect score: 2883  
Sequence: 1 MRDYDEVTAFLGEMGPFQRL.....HTRMUKDQGERPTLIKSTAF 557

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2883	100.0	557	2 JMW0089	organic cation tra
2	2496	86.6	557	2 JE0346	high-affinity carn
3	777	27.0	576	2 T22509	hypothetical prote
4	733	25.4	593	2 JC4884	organic cation tra
5	712	24.7	556	2 S50862	organic cation tra
6	589	20.4	794	2 T27870	hypothetical prote
7	547.5	19.0	745	2 T16565	hypothetical prote
8	522	18.1	527	2 T01019	transport protein
9	514	17.8	447	2 D8646	protein ZK455.8 [1
10	483	16.8	539	2 C96758	probable protein
11	470.5	16.3	518	2 B86299	hypothetical prote
12	411	14.3	528	2 T21682	hypothetical prote
13	387.5	13.4	751	2 C88485	protein F23P12.5 [
14	382.5	13.3	540	2 T25851	hypothetical prote
15	380.5	13.2	515	2 B96825	hypothetical prote
16	368.5	12.8	521	2 H86298	hypothetical prote
17	357	12.4	1222	2 C88504	protein B0361.3 [1
18	355.5	12.3	529	2 T23190	hypothetical prote
19	350	12.1	448	2 AC3602	transporter, mfs s
20	349	12.1	461	2 AE3208	mfs permease [impo
21	343	11.9	437	2 C97630	probable sugar tra
22	343	11.9	437	2 AB2854	mfs permease [impo
23	341.5	11.8	434	2 F75580	probable sugar tra
24	321.5	11.2	450	2 F95360	probable transmem
25	316	11.0	455	2 B83213	probable mfs trans
26	310	10.8	452	2 AD0300	probable transport
27	308	10.7	469	2 G65058	hypothetical prote
28	300.5	10.4	510	2 B88381	protein T22P7.1 [1
29	298.5	10.4	724	2 I50531	transmembrane tran

30	295.5	10.2	423	2 S74046	probable sugar tra
31	291.5	10.1	435	2 T15290	hypothetical prote
32	287	10.0	400	2 C69757	transporter homolo
33	287	10.0	422	2 G72234	hypothetical prote
34	279	9.7	524	2 T27082	hypothetical prote
35	277	9.6	448	2 B90498	permease, probable
36	274	9.5	683	2 S34961	synaptic vesicle p
37	272	9.4	478	2 T33985	hypothetical prote
38	271	9.4	480	2 T23608	hypothetical prote
39	267	9.3	443	2 E64725	yeau protein - Esc
40	267	9.3	443	2 H90634	probable transport
41	267	9.3	443	2 H85485	probable transport
42	266	9.2	423	2 T19030	hypothetical prote
43	265	9.2	459	2 G64937	hypothetical prote
44	265	9.2	459	2 D90939	probable transport
45	265	9.2	459	2 H85787	probable transport

## ALIGNMENTS

Query Match	100.0%; Score 2883; DB 2; Length 557;
Best Local Similarity	100.0%; Pred. No. 1.7e-212;
Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 MRDYDEVTAFLGEMGPFQRLIFPLLSASITIPNCFGLSSVFLLATPEHRCRPDAANLSS 60	
DB 1 MRDYDEVTAFLGEMGPFQRLIFPLLSASITIPNCFGLSSVFLLATPEHRCRPDAANLSS 60	
QY 61 AMRNHTVPLRLRDGRVPHSCRRYRLATTANFSAIGLEGRDVLGQLRQESCLDGMERS 120	
DB 61 AMRNHTVPLRLRDGRVPHSCRRYRLATTANFSAIGLEGRDVLGQLRQESCLDGMERS 120	
QY 121 QDYLSTIVTEWNLVCEDDMKAPLITSLFFVGLGSGFTSGQLSDRGRKNVLFVTMGQ 180	
DB 121 QDYLSTIVTEWNLVCEDDMKAPLITSLFFVGLGSGFTSGQLSDRGRKNVLFVTMGQ 180	
QY 181 TGSFSLQIFSKNEMFVFLVGMGOISNYAAFLVGLFEIILGKSVRIITSLFGCIFPA 240	
DB 181 TGSFSLQIFSKNEMFVFLVGMGOISNYAAFLVGLFEIILGKSVRIITSLFGCIFPA 240	
QY 241 FGMVPLFLFAFYTRDMRMLVALTMPGVLCAVMWFIIPSPRMLISQGFEEAEVIRKA 300	
DB 241 FGMVPLFLFAFYTRDMRMLVALTMPGVLCAVMWFIIPSPRMLISQGFEEAEVIRKA 300	
QY 301 AKANGIVVSTIPDPSLQDLSKKQOSHNIIDLRTWNIIMVTTWSIMLMTISVYFG 360	
DB 301 AKANGIVVSTIPDPSLQDLSKKQOSHNIIDLRTWNIIMVTTWSIMLMTISVYFG 360	
QY 361 LSLDPTNLHGDIIVNCFISAMVEVPAYVLAMLLQVLPRLRYSMATRALFGSSVILFMOLY 420	
DB 361 LSLDPTNLHGDIIVNCFISAMVEVPAYVLAMLLQVLPRLRYSMATRALFGSSVILFMOLY 420	
QY 421 PPDLYLATVLLVMGKFGVTAFAFSNVYVYTAELYPVTVBNMGVGSSTASRLGSLISPYF 480	

```
Db 421 PDLYLVAVLVAVGKFGVTAFAFSAFVYVYTAELPTVVRNMGVSSATASRLGSLSPF 480
Qy 481 VYLGAVDRPLPYILMGSLTILFAITLFLPESFGTLPDPTIDQMLRVGMRKRPSTHR 540
Db 481 VYLGAVDRPLPYILMGSLTILFAITLFLPESFGTLPDPTIDQMLRVGMRKRPSTHR 540
Qy 541 MKDGOERPTILKSTAF 557
Db 541 MKDGOERPTILKSTAF 557
```

## RESULT 2

JE0346  
high-affinity carnitine transporter, CT1 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: JE0346  
R:Sekine, T.; Kusuhara, H.; Utsunomiya-Tate, N.; Tsuda, M.; Sugiyama, Y.; Kanai, Y.; Endo  
Biochem. Biophys. Res. Commun. 251, 586-591, 1998  
A:Title: Molecular cloning and characterization of high-affinity carnitine transporter 4  
A:Reference number: JE0346; MUID:99011422  
A:Accession: JE0346  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-557 <SEQ>  
A:Cross-references: DDBJ:AB017260; NID:g3869208; PIDD:BAA34399.1; PIR:g3869209

```
Query Match 86.6%; Score 2496; DB 2; Length 557;
Best Local Similarity 85.3%; Pred. No. 6.2e-183;
Matches 475; Conservative 42; Mismatches 40; Indels 0; Gaps 0;

Qy 1 MDYDEVTAFLGEMGFQRLIFELLASITPENGFGSLSVFLIATPBRRCRVPAANLSS 60
Db 1 MDYDEVTAFLGEMGFQRLIFELLASITPENGFGSLSVFLIATPBRRCRVPAANLSS 60
Qy 61 AMRNHTVRLRDGRBVRHSRRRLATIANFSALGLEPGRDVLGQLQESCLDGEWFS 120
Db 61 AMRNHTVRLRDGRBVRHSRRRLATIANFSALGLEPGRDVLGQLQESCLDGEWFS 120
Qy 121 QDVYSTIVTEMNLVCEDDMKAPLTIISLFYGVLLGSFISGLSDRGRKNVLPVTGMQ 180
Db 121 QDVYSTIVTEMNLVCEDDMKAPLTIISLFYGVLLGSFISGLSDRGRKNVLPVTGMQ 180
Qy 121 KVEFLSTIVTEMNLVCEDDMKAPLTIISLFYGVLLGSFISGLSDRGRKNVLPVTGMQ 180
Db 121 KVEFLSTIVTEMNLVCEDDMKAPLTIISLFYGVLLGSFISGLSDRGRKNVLPVTGMQ 180
Qy 181 TGFSEQLTFSKNFEMFVVLVGVGQISNYVAFLVLTETLIGKSVRTIIFSTLGVCIYA 240
Db 181 TGFSEQLTFSKNFEMFVVLVGVGQISNYVAFLVLTETLIGKSVRTIIFSTLGVCIYA 240
Qy 241 EGYMVLPLFAFIRDMRLVALTMPGYLCVALMWFIPESPRMLISQGRPEAEYIIRKA 300
Db 241 EGYMVLPLFAFIRDMRLVALTMPGYLCVALMWFIPESPRMLISQGRPEAEYIIRKA 300
Qy 301 AKANGIVVPSITFDPSELQDLSKKQOSHNIIDLRTWNIIRVATIMSIMLMTISVGYG 360
Db 301 AKANGIVVPSITFDPSELQDLSKKQOSHNIIDLRTWNIIRVATIMSIMLMTISVGYG 360
Qy 361 ISLDPNHLGDIYVNCFLSAMVEPAVYLAAMLLOYLPRRYSMATALFLGSSVLLFMOLV 420
Db 361 ISLDPNHLGDIYVNCFLSAMVEPAVYLAAMLLOYLPRRYSMATALFLGSSVLLFMOLV 420
Qy 421 PPDLYLVAVLVAVGKFGVTAFAFSAFVYVYTAELPTVVRNMGVSSATASRLGSLSPF 480
Db 421 PPDLYLVAVLVAVGKFGVTAFAFSAFVYVYTAELPTVVRNMGVSSATASRLGSLSPF 480
Qy 481 VYLGAVDRPLPYILMGSLTILFAITLFLPESFGTLPDPTIDQMLRVGMRKRPSTHR 540
Db 481 VYLGAVDRPLPYILMGSLTILFAITLFLPESFGTLPDPTIDQMLRVGMRKRPSTHR 540
Qy 541 MKDGOERPTILKSTAF 557
Db 541 MKDGOERPTILKSTAF 557
```

```
RESULT 3
T22509
hypothetical protein F52F12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22509
R:Mathews, L.
submitted to the EMBL Data Library, December 1996
A:Reference number: 219573
A:Accession: T22509
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-576 <MIL>
A:Cross-references: EMBL:283228; PIDD:GAB05732.1; GSPDB:GN00019; CESP:F52F12.1
A:Experimental source: clone F52F12
C:Genetics:
A:Gene: CESP:F52F12.1
A:Map position: 1
A:Introns: 16/1; 23/3; 50/1; 80/3; 108/2; 134/3; 221/1; 257/2; 319/3; 358/2; 490/3; 5
```

Query Match 27.0%; Score 777; DB 2; Length 576;

Best Local Similarity 31.4%; Pred. No. 1.3e-51;

Matches 165; Conservative 118; Mismatches 225; Indels 18; Gaps 6;

```
Qy 3 DYDEVTAFLGEMGFQRLIFELLASITPENGFGSLSVFLIATPBRRCRVPAANLSS 61
Db 16 DFDEVLEQVNGVGYQIVPFELICPLSLPSAFSAFNPVGVGNPHTCHIEGKEYLNR 75
Qy 62 WRNHTVRLRDGRBVRHSRRRLATIANFSALGLEPGRDVLGQLQESCLDGEWFSQ 121
Db 76 LTNDTQIL-----SKQYNETQINVERFATSAP-VDYISDRISLVPCCNGMDYDN 124
Qy 122 DVYSTIVTEMNLVCEDDMKAPLTIISLFYGVLLGSFISGLSDRGRKNVLPVTGMQ 181
Db 125 STYDLSLVTERNLVCDQQAITEISTSFYSGTGNLFYVADKGRKRSFVILTVLI 184
Qy 182 GFSEQLTFSKNFEMFVVLVGVGQISNYVAFLVLTETLIGKSVRTIIFSTLGVCIYA 241
Db 185 VCGTASFAKDIESFILLRFTFGIAPALPQIFITCEPMGNSGR-IFSGLMTSLFGA 243
Qy 242 GYMVLPLFAFIRDMRLVALTMPGYLCVALMWFIPESPRMLISQGRPEAEYIIRKA 301
Db 244 AMALLGVVAMFIRMRQLTFECNAPFAFYIIFYFLPESPRMSVSGKADAKOLKKA 303
Qy 302 KANG---IVPSTFDPSELQDLSKKQ--QSHNIIDLRTWNIIRVATIMSIMLMTISV 356
Db 304 KMGKSNVDVDELVDLSKKNQNAEKEKRSNINVDLFTPLRKRKTLIVYIWMNAI 363
Qy 357 GYFGLSLDTPNLHGDIFVNCFLSAMVEPAVYLAAMLLOYLPRRYSMATALFLGSSVLLF 416
Db 364 IYNGLTNLVNSLPPVDVWSFILINGAVELPGYFVWPLPDQAGRRRLATIMVCGIGCS 423
Qy 417 MQLVPPDLYLVAVLVAVGKFGVTAFAFSAFVYVYTAELPTVVRNMGVSSATASRLGSL 476
Db 424 AMFMPDGYPWLVAASASTIGFVGSGFAYIYIRAGELIYIPVVAIAGMSSMVGSGLL 483
Qy 477 SPFYVYLGAVDRPLPYILMGSLTILFAITLFLPESFGTLPDPTID 522
Db 484 APHTVNGIKVILKILPLLIMGLMALSAGILTFPLPETLIGALPMTIE 529
```

RESULT 4  
JC4884  
organic cation transporter protein 2 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 10-Sep-1996 #sequence\_revision 18-Oct-1996 #text\_change 05-Nov-1999  
C:Accession: JC4884  
R:Okuda, M.; Saito, H.; Urakami, Y.; Takano, M.; Inui, K.  
Biochem. Biophys. Res. Commun. 224, 500-507, 1996  
A:Title: cDNA cloning and functional expression of a novel rat kidney organic cation  
A:Reference number: JC4884; MUID:96295517

A:Accession: J04884  
 A:Molecule type: mRNA  
 A:Residues: 1-593 <ORF>  
 A:Cross-references: DDBJ:D83044; NID:91502282; PIDN:BA11754.1; PID:d1012421; PID:915022  
 A:Experimental source: kidney  
 C:Comment: This protein is responsible for the transport of cationic drugs in kidney.

Query Match 25.4%; Score 733; DB 2; Length 593;  
 Best Local Similarity 35.2%; Pred. No. 3,1e-48;  
 Matches 198; Conservative 89; Mismatches 234; Indels 42; Gaps 15;

QY 1 MRDYEVTAFLGEMGFQRLIFLLSASIIIPNGFTGL--SSVFLLATPEHRCRPDAA NL 58  
 DB 1 MSTVDLLIEHGEFLPFQKQTFPLL--ALLSGAFPIIVGVIFGLTFPHDHCHMGSGAKL 58  
 QY 59 SS--AMR-----NHTVP--LRLRDGEVPHSCRRR-----LATIANFSLGLEPCRDV 103  
 DB 59 SORCGMSQAEELNYTPVGLGPSDEASFLSQCMRYEVDNOSTLDCVPLSLA-----A 112  
 QY 104 DLGQLEQSCLDGMEFSQDYLSTVTEMNIVCEDDKAPLTLISLFFVGLSGFISQGL 163  
 DB 113 DRNQLPCLPCEHGWYNTPG--SSIVTEFNLVCAHSMMLDLPQSVVNGFTIGAMMIGYL 170  
 QY 164 SDRGRKNVLEVTMGQTFSEFLQIFSKNFEFVLEVLVGMQGISNVAAVFLGTETIG 223  
 DB 171 ADRFGKRCCLLVTLINISGALMAISPVAMLVFRLOGLVSKAGLIGYIILTEFVG 230  
 QY 224 KSVRIEFTLGVCFIYAR--GYMVLPLFAFIRDMRMILVALTMRGVLCVALMWFIPESP 281  
 DB 231 LGYR---RMVGICVQIATVGLILLIAGVAVYIPMWRMLQFVLTLPNCFLLYFCIDESP 287  
 QY 282 RLLISQGRFEAEVYIIRKAKANGIVVSTIFDPSLQDSSKKQOQSNIIIDLRNMIR 341  
 DB 288 RFLISQNKIIVAKMTIKIIRAKNGKSVSVSLQNTLPDDEACKKLKPS--IIDLVTPTQIR 345  
 QY 342 MYTWSIMLMNTISVYFGSLDTPENLHGD--IFVNCFLSAMVEPAAVYLA ML LQYLP RR 400  
 DB 346 KHTLIMNMFSTSVLYOGLIMHM--GLAGDNILDFEFSALVEPPAFIILITLDRGR 404  
 QY 401 YSMATALLGGSVLLFMQVLPDLYLATVLYVMGKEGVTAESMNVYTTAE LPTVYRN 460  
 DB 405 YPMASNNVAGAACIASYFIPDDLQWLKITTACLRMGITMAYEMVCILVNAELPTYIRN 464  
 QY 461 MGVGVSTASRLSITSPFYV--LGAYDRFLPYILMGSLLTIFALTLTLPESGTPLPD 519  
 DB 465 LGVLVSSMCIDIGLITPFLYRLFDIMMERPLVYFAVGVLAGALVLLLPETGKALPE 524  
 QY 520 TID-----QMLRVKGMKHKRTPS 537  
 DB 525 TIEDAENMQRRKKERKENLPSS 547

RESULT 5  
 S50862  
 organic cation transport protein OCT1 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 05-Nov-1999  
 C:Accession: S50862; S78533; I58089  
 R:Grendemann, D.; Gorboulev, V.; Gambaryan, S.; Veyhl, M.; Koepsell, H.  
 Nature 372, 549-552, 1994  
 A>Title: Drug excretion mediated by a new prototype of polyspecific transporter.  
 A:Reference number: I58089; MUID:95082907  
 A:Accession: S50862  
 A:Molecule type: mRNA  
 A:Residues: 1-556 <GRU>  
 A:Cross-references: EMBL:X78855; NID:9633621; PIDN:CAA55411.1; PID:9633622  
 A:Experimental source: kidney  
 R:Gorboulev, V.G.  
 submitted to the EMBL Data Library, January 1995  
 A:Accession: S78533  
 A:Molecule type: mRNA

A:Residues: 1-342, 'N', 344-556 <GOR>  
 A:Cross-references: EMBL:X78855; NID:9633621; PIDN:CAA55411.1; PID:9633622  
 C:Keywords: glycoprotein; phosphoprotein; transmembrane protein  
 F;20-46/Domain: transmembrane #status predicted <TM1>  
 F;154-171/Domain: transmembrane #status predicted <TM2>  
 F;178-197/Domain: transmembrane #status predicted <TM3>  
 F;243-260/Domain: transmembrane #status predicted <TM4>  
 F;267-283/Domain: transmembrane #status predicted <TM5>  
 F;350-366/Domain: transmembrane #status predicted <TM6>  
 F;360-398/Domain: transmembrane #status predicted <TM7>  
 F;406-425/Domain: transmembrane #status predicted <TM8>  
 F;435-452/Domain: transmembrane #status predicted <TM9>  
 F;469-485/Domain: transmembrane #status predicted <TM10>  
 F;484-514/Domain: transmembrane #status predicted <TM11>  
 F;71,97,113,432/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F;286,492/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pred  
 F;296,343,550/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status

Query Match 24.7%; Score 712; DB 2; Length 556;  
 Best Local Similarity 33.8%; Pred. No. 1.2e-46;  
 Matches 193; Conservative 96; Mismatches 230; Indels 52; Gaps 14;

QY 1 MRDYEVTAFLGEMGFQRLIFFL--SASIIIPNGFTGISVFLATPEHRCRPDAA NL 57  
 DB 1 MPTVDVLEQVGFQFQKQTFPLLCLISASLAP--YVGI--VFLGFTGHQCNQVGE 57  
 QY 58 LSS--AMR-----NHTVP--LRLRDGEVPHSCRRRL-----ATIANFSL 95  
 DB 58 LSORCGMSQAEELNYTPVGLGPSDEASFLSQCMRYEVDNOSTLDCVPLSLIVANRSGL 117  
 QY 96 GLPGRVVDLGQLEQSCLDGMEFSQDYLSTVTEMNIVCEDDKAPLTLISLFFVGL 155  
 DB 118 PLGP-----CEHGWY--DTPGSSIVTEFNLVCGDAMKVDLFGQCVNLGFTL 162  
 QY 156 GSFISQSDRGRKNVLEVTMGQTFSEFLQIFSKNFEFVLEVLVGMQGISNVAAV 215  
 DB 163 GSLVGYIADRFRKCLLVTLTVTSVSGVLTAVADYISMFLRLQGMVSKGSVSGY 222  
 QY 216 VLGTETIGKSVRIEFTLGVCFIYAR--GYMVLPLFAFIRDMRMILVALTMRGVLCVAL 273  
 DB 223 TLITTEFGSGYR--RTTALIXOMAFVGLVAGVAVYIPDMWRMLQVLSPTFLFLY 279  
 QY 274 MWFIPESPRLISQGRFEAEVYIIRKAKANGIVVSTIFDPSLQDSSKKQOQSNIIID 333  
 DB 280 YMFVPESPRMLISQKTRTAVRIMEQIAQRNGKVPADLKMCLTEDDASEKR--SPSFAD 337  
 QY 334 LRLTNIRMTYIMSIMLMNTISVYFGSLDTPENLHGDIFVNCFLSAMVEPAAVYLA ML 393  
 DB 338 LFRPTLRKHTVILMYLMFSCAVLYOGLIMHVGATGANLYDFEYSLSLVEPPAFIILVT 397  
 QY 394 LQYLP RRYSMATALLGGSVLLFMQVLPDLYLATVLYVMGKEGVTAESMNVYTTAE L 453  
 DB 398 IDRIGRIYPTAASNLVGAACLIMIFIPHELMVNTYTLACLRMGATIVLQMWCLVNAEL 457  
 QY 454 YPTVVRNMGVGVSTASRLSITSPFYV--LGAYDRFLPYILMGSLLTIFALTLTLPES 512  
 DB 458 YPFIIRNLGMVVSALCDLGGITPFRMVEWQALPILFGVGLTAGAMTLLPPT 517  
 QY 513 FGTPLEPDTIDMLRVKGMKHKRTPSHTRMUK 543  
 DB 518 KGVALPETIEAENL--GRKSKAKENTYIQ 547

RESULT 6  
 T27870  
 hypothetical protein ZK455.8 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T27870  
 R:White, S.  
 submitted to the EMBL Data Library, November 1995  
 A:Reference number: Z20432

A:Accession: T27870  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-794 <ML>  
A:Cross-references: EMBL:Z66567; PIDN:CAA91492.2; GSPDB:GN00028; CESP:ZK455.8  
A:Experimental source: clone ZK455  
C:Genetics:  
A:Gene: CESP:ZK455.8  
A:Map position: X  
A:Introns: 52/3; 119/2; 158/1; 207/3; 291/3; 335/2; 428/3; 610/2; 636/3; 733/3

Query Match  
Best Local Similarity 27.8%; Pred. No. 4.3e-37;  
Matches 167; Conservative 113; Mismatches 231; Indels 90; Gaps 22;

3 DYDE---VTAFLGEMGPORLIFLLASIIIPNGFTGLSSVFLIATPEHRCRP---D 54  
155 DPEEDDLGILQIGGCSYQIIVLISVQVPHAMFNLSVYMYMPDWMCKIPFNDE 214  
55 AANLSAMRNHT-----VPLRLRDGEPH--SCRRIATIANFSAIGLEPRD 102  
215 SESAEIGYNTYWDOLNSTIAFPRTFNKRNELHHDCHFE---RDYVHIKISPAQ 270  
103 V-DLGLEQESCLDGEFESODVYLSITVTEBNLYCEDDKAPLTSLFFVGLGSFTSG 161  
271 VDMNATGINKCKEWEYTSMDRTIYEMNRVCDNNKRAHMYSLGLVGCIFYG 330  
162 QLSDFGRKNVLFVTGMOTG--FSLQIFSKNEFM--VLEVLVGMGQISNYAAFLGT 219  
331 FTSDFGRKTAI-TGFGILTMLFGLLTYSKFEIIFLVRFLLAATNAD--LAAYVLCM 388  
220 ELIGKSVRIIFSTGICIFYAGYVLPFAFIFRDMRLIVALLMPGYLCALMWFIPE 279  
389 EYGTGYRSIVGSL-TQAFWAGCIAPLALIALTSMTWILHLCVLIHLSIMLFLPE 447  
280 SPRWLISQGRFEAEVILIRKAKANGIVPSTIFPSELQDSSKQOOSH---ILDLL 335  
448 SPRWLILNKKTKQAEKIIIEACHYKNSRLPS---DLGLVRAHEKKMKMKHNEKPEYHLE 504  
336 RFWNIAMVTIMSLMM-----TISVY-----FELSLDT 365  
505 RSESLFRNVVLEIYVWIDKNIFONSRFQYSDSNCSGLRNQCYIVRSVITRKVRLTSE 564  
366 P-----NLHGDIVNCFLSAMVEPAPVYVLAAILLOLXLPYRSMAATLFGGSVLT 414  
565 KFYRNKLCQFRVFDGNFELNNAMAGATLPLVFCVPLLR-MGKRKRSQMLVF--GSGLF 622  
415 LFMQVLPDLVY---LATVLVWGKFGVTAFAFNVYVYATVLYPVVRRNGGVSSTA 469  
623 LITSVY---MYRKOSTLALIFMLLSKACIOGSFNILYFTSELPTVVRNSAVGSSVY 679  
470 SRLGSLSPFYVIGAYDR-FLPYILMGSLLTILTLLEFPESSGFTPLDIT--QOMLRV 527  
680 ARGAGASGITALISDVTMPLVPMITIFACFSILAGCVLLLEPOTGLPLDITLDSVQNV 739

528 K 528  
740 K 740

RESULT 7  
T16565  
hypothetical protein K05F1.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16565  
R:Wohldmann, P.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of C. elegans cosmid K05F1.  
A:Reference number: Z18537  
A:Accession: T16565  
A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA  
A:Residues: 1-745 <MO>  
A:Cross-references: EMBL:029377; NID:9868173; PID:9868176; PIDN:AAA68713.1; CESP:K05F  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:K05F1.6  
A:Introns: 70/2; 160/3; 205/3; 249/3; 293/2; 350/1; 386/3; 462/3; 546/2; 669/3; 680/3

Query Match  
Best Local Similarity 19.0%; Score 547.5; DB 2; Length 745;  
Matches 165; Conservative 105; Mismatches 243; Indels 87; Gaps 17;

3 DYDEVTAFLGEMGPORLIFLLASIIIPNGFTGLSSVFLIATPEHRCRP-----D 53  
112 DFEGLIITIGCCRWQWQIIVYVLAQIIPAMFNLSVYMYMPDWMCKIPFNDE 171  
54 -----DANLSAMRNHTVPLRLRDGEPHSC-----RRIATIANFSA 94  
172 ISSHGPYLMGVDDIKNISVFPNAN-----SDGAYQRDSCYFERSEERYROLRMPLET 226  
95 LGLEPRDVIDLGLEQESCLDGEFESODVYLSITVTEBNLYCEDDKAPLTSLFF-VGY 153  
227 AMSEAMNDV---APKRCQ-QAYHEKDVWETIYVDNLVC--DSMFAGHAMFYSIGY 280  
154 LIGSEFISGOLSDRGRKNVLFVTGMOTGFSFIQI-----FSKNEFMVLEVLVGMG 206  
281 LIGCVLGSIASDKIGRKPTII-----GFGILSMGLVFLPNNDYPMFLIRLSAIC 333  
207 QISNYAAFLVGLTEILKSVRIIFSTGICIFYAGYVLPFAFIFRDMRLIVALLTMP 266  
334 NEADLAAYTLCEWITGTYKRAMGSM-LQAPMGALLALIALYLRKSKTIQVIAAGL 392  
267 GVLGVALLMWFIPEPRRLISQGRFEAEVILIRKAKANGIVSTIFD-----PS 316  
393 HEMSIIFTCSTIPESPRLMWONRVSEAEVIRACRE-----PPEPNMCTSKCGLPS 447  
317 ELQDLSKKQOOSH-----NIIDLRTWNIIRWYTIMSLMMTISVYGLSLU-----D 364  
448 DLELVSHREKLNKNGKIGFDLFTMKELRYRTISCVIPFATMLVYVGLWALSDQS 507  
365 TPN---LHGDIVNCFLSAMVEPAPVYVLAAILLOLXLPYRSMAATLFGGSVLLFMQVLP 421  
508 APGRITLTFGYFHLNNGIAGIEIPLTFACVWMMQIGRKKMLTILTISGIFIVAMLSVY 567  
422 PDLVYLAFTVLVWGKFGVTAFAFNVYVYATVLYPVVRRNGGVSSASTLGSILSPYEV 481  
568 SGHYMLALAFMYFGKIAVQCAFENILYFTSELPTVVRNNAVGTSMVAFSGSLSYIA 627  
482 YLGAYD-RFLPYILMGSLLTILTLLEFPESSGFTPLDITQMLR-VKGMKRRKTPSHT 539  
628 LLSNISLPIYPMIIFAVFSIFAGMLVFLPETSERKPLPETLIDALINLEPTKQFKTTEST 687

RESULT 8  
T01019  
transport protein homolog YUP8H12R.2 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 22-Oct-1999  
C:Accession: T01019  
R:Theologis, A.; Vysotskaia, V.S.; Osborne, B.I.; Schwartz, J.R.; Federspiel, N.A.; K  
oelner, P.; Davis, R.W.  
submitted to the EMBL Data Library, May 1998  
A:Description: Arabidopsis thaliana chromosome 1 YAC YUP8H12R sequence.  
A:Reference number: Z14227  
A:Accession: T01019  
A:Status: translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-527 <THE>  
A:Cross-references: EMBL:AC002986; NID:92494106; PID:93152576; GSPDB:GN00059; ATSP:YU  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Gene: ATSP:YUP8H12R.2





QY 374 VNCFLSMVEVPAYVLAMLLQYLPRLRYSMATALFLGSVLLPMQVLPDLYLATVLM 433  
 Db 355 INALIMS-----LKWVL-----NIFELADLKRKLGRKLIHLISKLITF 395  
 QY 434 VGRFV-----TAARSMVYVTA---ELYPVVRNMGVGS 466  
 Db 396 ICVESIFVYFEYEDYSILIRVATLLASATASQVFTKSMVLMFEYPIVRISAVSF 455  
 QY 467 STARLGSILSPFYVLGADRPLPYLMSLITLTAI-LTLPLEPSFGPLPDTIQ 523  
 Db 456 SSASRIGHTLPOLFLICPY-KSLPYALITGFLDAIAQLRLPRTKGPALPETMPE 512

## RESULT 13

C88485  
 Protein F23P12.5 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-May-2001 #sequence, revision 10-May-2001 #text-change 10-May-2001

C:Accession: C88485

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_eleg

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: C88485

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-751 <STO>

A:Cross-references: GB:chr\_III; PIDN:AAA20607.1; PID:9529214; GSPDB:GN00021; CESP:F23P12

C:Genetics:

A:Gene: F23P12.5

A:Map position: 3

Query Match 13.4%; Score 387.5; DB 2; Length 751;  
 Best Local Similarity 25.8%; Pred. No. 1e-21;  
 Matches 106; Conservative 89; Mismatches 189; Indels 27; Gaps 9;

QY 127 TIVTWMNVLC-EDMKAPLITSLFVGVLLGSGISGQSDRFGRKNVLFVTMGMOGTSF 185  
 Db 303 SMVDFKMFCCGKAYDAWVAITQIFGLIATYGGHGHGKRPVSFGISVGLIFGV 362  
 QY 186 LQIFSKNEMFVFLVLMGQISNVVAFLGTEILKSVRIIFSTLGVCFYAFGY-- 243  
 Db 363 ASGFAPSEVFAAFRFYIGTSTSLIYFYVILEFIEPQGVPLRS-----FFNWGAR 417  
 QY 244 MVLPLFAVFIIDWRMLVALTMPGVLCVALLMWFIPESPRLTISQREFEAEVIRKAKA 303  
 Db 418 LVFTLACFCGWRSAATSLSLPILPVLLIPESPKWNTKKRFDAARAEKRVAML 477  
 QY 304 NGIIVPSTIPPELOD-LSSKKOOSHNIIDLRTWNTRMVIMSIMAMNISGVYELS 362  
 Db 478 SGIPYVNDSDSIEITSEKLEEKSTIYTKMDLFTSWTAYRTIVVGSLSAFESD 537  
 QY 363 IDTPNLHGDIYVNCFLSMAVEVPAYVLAMLLQYLP---RYSM--ATALFLGGSVLLF 416  
 Db 538 LNSGMLAGNFYLSQVSGAVTAFAKIFPFLDTPVPSDRRLHQYPLAMILCVIMV 597  
 QY 417 MOLVPP-----DLYLATVLVMGKFGVTAAFSMVYITAEIYTVVRNMGVSS 467  
 Db 598 LMLPESDGSOGSRDLAI--IINIGVSFEITWADCYAAVACFPKTRITIGTICS 655  
 QY 468 TMSRGLSTSPFYVLGADRPLPYLMSLITLTAI-LTLPLEPSFGPLPDTIQ 517  
 Db 656 LIARGTALAPQAMATLSDIRPAPAVVCSITGTLISLCVFLPDKGYDL 706

## RESULT 14

hypothetical protein T01B11.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence, revision 15-Oct-1999 #text-change 15-Oct-1999

C:Accession: T25851  
 R:Geisel, C.; Steillyes, L.  
 submitted to the EMBL Data Library, December 1996  
 A:Description: The sequence of C. elegans cosmid T01B11.  
 A:Reference number: Z20099  
 A:Accession: T25851  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-540 <GET>  
 A:Cross-references: EMBL:U80931; PIDN:AAB38002.1; GSPDB:GN00022; CESP:T01B11.5  
 A:Experimental source: strain Bristol N2; clone T01B11  
 C:Genetics:  
 A:Gene: CESP:T01B11.5  
 A:Map position: 4  
 A:Introns: 85/1; 128/3; 209/2; 269/2; 352/2; 381/3; 489/3

Query Match 13.3%; Score 382.5; DB 2; Length 540;  
 Best Local Similarity 26.6%; Pred. No. 1.6e-21;  
 Matches 120; Conservative 78; Mismatches 178; Indels 75; Gaps 14;

QY 127 TIVTWMNVCEDEWKAPLITSLFVGVLLGSGISGQSDRFGRKNVLFVTMGMOGTSF 186  
 Db 124 SYOERNYICEDAKKKNNTIYQTFGVLAGAIFGVSDTFGRKRALISTGNGLFMT 183  
 QY 187 QIFSKNEMFVFLVLMGQISNVVAFLGTEILKSVRIIFSTLGVCFYAFGYVVL 246  
 Db 184 TAVSPDLFTYMWRTLAGVFTGVTVVQVWFMENIPRDRMIQV---SITWSPLLIF 240  
 QY 247 PLFAYFIIDWRMLVALTMPGVLCVALLMWFIPESPRLTISQREFEAE--YIRKAKA 303  
 Db 241 PFVAWLQDWRMVSVAIASSIATFLACILLESPPRLIÖKRIEARSLSLRIRKTD- 299  
 QY 304 NGIIVPSTIPD-----PSELQDSSKKOOSHNIIDLRTWNT--RMVITMSI 348  
 Db 300 -----LYDEFERKOLDEVLHVEAEKHAARSSKKAKKTFIIFLFTWMLAQSLFTIG 351  
 QY 349 LMMWTISVGYEGLSLDTPNLHGDIYVNC-----FLSMVEVPAYVLAMLLQYLP 400  
 Db 352 ITCTFTIV--YSLMNYNEKLSGLYNNLAIMGASRMINIAVSADYKLPMFGKRMV-NO 408  
 QY 401 YSNATALFLGGSVLLFMQVLPDLY--YLATVLMVGKFGVTAAFSMV-----YVYTA 451  
 Db 409 IAMCTLFA-----LMTAVVLYFGYNGQLTAI---GTVAAVSCSQLFATKYMVN 457  
 QY 452 ELVPTVVRNMGVSSPSFASRLSGISLSPFYVLGAYDRPLPYLMSLITLT-AITLFLP 510  
 Db 458 ELVPTAVRNLAVASVSTMSRIGSMFQFLYLDYAEWIPYAVLFCQLXDFIVLSLFLP 517  
 QY 511 ESFGTPLPDTIDQMLRVKGMKHKRTPSHTPM 541  
 Db 518 ETKGVIL-----ENHLPRKHRI 535

## RESULT 15

hypothetical protein T8K14.17 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence, revision 02-Mar-2001 #text-change 31-Mar-2001

C:Accession: B96825

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.K.; Hultzer, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: B96825

A>Status: preliminary





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2002, 02:15:01 ; Search time 39.73 Seconds

(Without alignments)  
542.834 Million cell updates/sec

Title: US-09-521-195b-3

Perfect score: 2883

Sequence: 1 MRDYDEVTAFLGEMGFQRL.....HTRMLKDGQERPTTIKSTAF 557

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2883	100.0	557	1	OCN2_HUMAN
2	2513	87.2	557	1	OCN2_MOUSE
3	2509	87.0	557	1	OCN2_RAT
4	387.5	13.4	751	1	YLX_CABEL
5	357	12.4	1222	1	YMP3_CABEL
6	355.5	12.3	529	1	YOUN_CABEL
7	308	10.7	445	1	YGCS_ECOLI
8	291.5	10.1	435	1	YTI3_CABEL
9	287	10.0	400	1	YCE1_BACSU
10	267	9.3	443	1	YAAO_ECOLI
11	265	9.2	459	1	YDUR_ECOLI
12	262.5	9.1	742	1	SYV2_RAT
13	256	8.9	482	1	YFIC_BACSU
14	255.5	8.9	452	1	YDUE_ECOLI
15	252.5	8.8	457	1	PCAR_ACICA
16	250.5	8.7	472	1	ARAE_KLEOX
17	243.5	8.4	551	1	HGT1_KLULA
18	241.5	8.4	472	1	ARAE_ECOLI
19	240.5	8.3	495	1	GTR3_CANFA
20	239.5	8.3	451	1	YTAI_CANFA
21	239	8.3	461	1	YBAC_BACSU
22	237.5	8.1	496	1	GTR3_HUMAN
23	234	8.1	494	1	GTR3_SHEEP
24	230	8.0	494	1	GTR3_BOVIN
25	227	7.9	490	1	GTRI_CHICK
26	226.5	7.9	464	1	GALP_ECOLI
27	224.5	7.8	413	1	MUCK_ACICA
28	223	7.7	451	1	GTRI_PIG
29	222	7.7	763	1	RGT2_YEAST
30	221	7.7	466	1	BENK_ACICA
31	221	7.6	492	1	GTRI_RAT
32	220	7.6	492	1	GTRI_BOVIN
33	219.5	7.6	491	1	XYLE_ECOLI

34	219	7.6	493	1	GTR3_MOUSE	P32037 mus musculus
35	217	7.5	492	1	GTRI_MOUSE	P17809 mus musculus
36	217	7.5	496	1	GTR3_CHICK	P28568 gallus galli
37	216	7.5	492	1	GTRI_HUMAN	P11166 homo sapien
38	215	7.5	492	1	GTRI_RABIT	P13355 oryctolagus
39	213.5	7.4	486	1	YGR4_YEAST	P53142 saccharomyc
40	213	7.4	448	1	PCAR_PSEPU	Q51955 pseudomonas
41	213	7.4	481	1	SNF3_YEAST	P10870 saccharomyc
42	208	7.2	493	1	GTR3_RAT	Q07647 rattus norv
43	207.5	7.2	567	1	HXR9_YEAST	P40885 saccharomyc
44	206.5	7.2	567	1	HXTA_YEAST	P54862 saccharomyc
45	204	7.1	547	1	GTRI_LEIDO	Q01440 leishmania

#### ALIGNMENTS

RESULT 1  
ID OCN2\_HUMAN STRAND; PRT; 557 AA.  
AC 076082;  
DT 16-OCT-2001 (rel. 40, Created)  
DT 16-OCT-2001 (rel. 40, Last sequence update)  
DE Organic cation/carnitine transporter 2 (Solute carrier family 22, member 5) (High-affinity sodium-dependent carnitine cotransporter).  
GN SLC22A5 OR OCTN2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-98289574; PubMed-9618255;  
RA Wu X., Prasad P.D., Leibach F.H., Ganapathy V.;  
RT "cDNA sequence, transport function, and genomic organization of human OCTN2, a new member of the organic cation transporter family.";  
RL Biochem. Biophys. Res. Commun. 246:589-595(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE-98352077; PubMed-9685390;  
RA Tamai I., Ohashi R., Nezu J.-I., Yabuuchi H., Oku A., Shimane M., Sai Y., Tsuji A.;  
RT "Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2.";  
RL J. Biol. Chem. 273:20378-20382(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99113835; PubMed-9916797;  
RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N., Nikaido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsubhi T., Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M., Tsuji A.;  
RT "Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter.";  
RL Nat. Genet. 21:91-94(1999).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE-99384224; PubMed-10454528;  
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H., Chen J., Conway S.J., Ganapathy V.;  
RT "Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic cation/carnitine transporter.";  
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).  
RN [5]  
RP VARIANT CDSP GIN-169.  
RX MEDLINE-99355597; PubMed-10425211;  
RA Burdink B., Kreuder J., Schweitzer S., Vorgerd M., Gempel K., Gerlitz K.-D., Killmann M.W.;  
RT "Carnitine transporter OCTN2 mutations in systemic primary carnitine deficiency: a novel Arg169Gln mutation and a recurrent Arg282Ser

RT mutation associated with an unconventional splicing abnormality.";  
RL Biochem. Biophys. Res. Commun. 261:484-487(1999).  
[6]  
RN VARIANT CDSF CYS-211.  
RP MEDLINE=99408248; PubMed=10480371;  
RX Vaz F.M., Scholte H.R., Ruiter J., Hussaerts-Odiik L.M.,  
RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,  
RA Wanders R.J.A.,  
RT "Identification of two novel mutations in OCTN2 of three patients with  
RT systemic carnitine deficiency.";  
RL Hum. Genet. 105:157-161(1999).  
[7]  
RN VARIANT CDSF LEU-478.  
RP MEDLINE=99172075; PubMed=10072434;  
RX Tang N.L., Ganapathy V., Wu X., Hui J., Seth P., Yuen P.M.,  
RA Wanders R.J., Fox T.F., Hjelm N.M.,  
RT "Mutations of OCTN2, an organic cation/carnitine transporter, lead to  
RT deficient cellular carnitine uptake in primary carnitine deficiency.";  
RL Hum. Mol. Genet. 8:655-660(1999).  
[8]  
RN CHARACTERIZATION OF VARIANT CDSF LEU-478, AND MUTAGENESIS.  
RP MEDLINE=2002865; PubMed=10559218;  
RX Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;  
RT "Mutations in novel organic cation transporter (OCTN2), an organic  
RT cation/carnitine transporter, with differential effects on the  
RT organic cation transport function and the carnitine transport  
RT function.";  
RL J. Biol. Chem. 274:33388-33392(1999).  
[9]  
RN VARIANT CDSF ARG-283 AND PHE-446.  
RP MEDLINE=20081068; PubMed=10612840;  
RX Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,  
RA Tsuji A.;  
RT "Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a  
RT patient with primary systemic carnitine deficiency.";  
RL Hum. Mutat. 15:118-118(2000).  
[10]  
RN VARIANT CDSF LYS-452.  
RP MEDLINE=20145665; PubMed=10679939;  
RX Wang Y., Kelly M.A., Cowan T.M., Longo N.;  
RT "A missense mutation in the OCTN2 gene associated with residual  
RT carnitine transport activity.";  
RL Hum. Mutat. 15:238-245(2000).  
[11]  
CC -I- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE  
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE  
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF  
CC CARNITINE.  
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -I- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,  
CC HEART AND PLACENTA.  
CC -I- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY  
CC CARNITINE DEFICIENCY (CDSF). CDSF IS AN AUTOSOMAL RECESSIVE  
CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE  
CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND  
CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL  
CC MYOPATHY OR CARDIOMYOPATHY.  
CC -I- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
CC CATION SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
CC EMBL: AF057164; AAC24828.1; -  
DR EMBL: AB015050; BAA29023.1; -  
DR EMBL: AB016625; BAA36712.1; -  
DR MIM: 603377; -  
DR MIM: 212140; -  
DR InterPro: IPR003362; sub\_transporter.

DR	pfam; PF00083; sugar_tr; 1.
DR	PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
KW	TransPort; Transmembrane; Glycoprotein; Disease mutation.
FT	TRANSMEM 21 41 POTENTIAL.
FT	TRANSMEM 143 163 POTENTIAL.
FT	TRANSMEM 173 193 POTENTIAL.
FT	TRANSMEM 196 216 POTENTIAL.
FT	TRANSMEM 233 253 POTENTIAL.
FT	TRANSMEM 258 278 POTENTIAL.
FT	TRANSMEM 343 363 POTENTIAL.
FT	TRANSMEM 372 392 POTENTIAL.
FT	TRANSMEM 414 434 POTENTIAL.
FT	TRANSMEM 437 457 POTENTIAL.
FT	TRANSMEM 489 509 POTENTIAL.
FT	CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT 169 169 R -> Q (IN CDSP). /FTID-VAR_009252.
FT	VARIANT 211 211 Y -> C (IN CDSP). /FTID-VAR_009253.
FT	VARIANT 283 283 W -> R (IN CDSP). /FTID-VAR_009254.
FT	VARIANT 446 446 V -> F (IN CDSP). /FTID-VAR_009255.
FT	VARIANT 452 452 E -> K (IN CDSP). /FTID-VAR_009256.
FT	VARIANT 478 478 P -> L (IN CDSP); LOSS OF CARNITINE TRANSPORT BUT STIMULATED ORGANIC CATION TRANSPORT. /FTID-VAR_009257.
FT	MUTAGEN 352 352 M->R: LOSS OF BOTH CARNITINATE AND ORGANIC CATION TRANSPORT FUNCTIONALITIES. 928B1E6FEF63C48D CRC64;
SO	SEQUENCE 557 AA; 62751 MW; 928B1E6FEF63C48D CRC64;

Query Match	Similarity	100.0%	Score 2883	DB 1	Length 557
Local Similarity	100.0%	Pred. No. 4,6e-177			
Matches 557	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
OY	1	MNDYEVATAFEGMGPFORLIFELLASITIPGFGSSVFLIATPERRRCRPDANLSS	60		
Db	1	MNDYEVATAFEGMGPFORLIFELLASITIPGFGSSVFLIATPERRRCRPDANLSS	60		
OY	61	AMRNHTVPLRLDGRGVPHSCRRYRLATIANFSALGLEGRDVGDLGQLEQESCLDGMERS	120		
Db	61	AMRNHTVPLRLDGRGVPHSCRRYRLATIANFSALGLEGRDVGDLGQLEQESCLDGMERS	120		
OY	121	QDVIYSTIVTEBNLVACEDDKAPLTIISLFPYGVLLGSFTSGQLSRFGKKNVLPYTMQ	180		
Db	121	QDVIYSTIVTEBNLVACEDDKAPLTIISLFPYGVLLGSFTSGQLSRFGKKNVLPYTMQ	180		
OY	181	TGEFSLQIFSKNEFVVLFTVVGGOISNYAAVLGTEILIGKSVRIEFTLGICIFYA	240		
Db	181	TGEFSLQIFSKNEFVVLFTVVGGOISNYAAVLGTEILIGKSVRIEFTLGICIFYA	240		
OY	241	FGYANVLPLFAYIRIMRMILVALITMPGVLCVLAHMFIESPFWLLISQGFEEAEVYIIRKA	300		
Db	241	FGYANVLPLFAYIRIMRMILVALITMPGVLCVLAHMFIESPFWLLISQGFEEAEVYIIRKA	300		
OY	301	AKANGIIVPSTIEPDESELQDSSKKQOOSHNIIDLRTWNIRMTIMSLMWTISVGYFG	360		
Db	301	AKANGIIVPSTIEPDESELQDSSKKQOOSHNIIDLRTWNIRMTIMSLMWTISVGYFG	360		
OY	361	LSLDTPNLHGLDFVNVCFISAMYEYAVYLAWLLOYLPRRYSMATALFLGGSVLLFMQLY	420		
Db	361	LSLDTPNLHGLDFVNVCFISAMYEYAVYLAWLLOYLPRRYSMATALFLGGSVLLFMQLY	420		
OY	421	PDDLTYLATVLMVCKFGYTAFSWVYYTALYPTVYNNMGVGSSTASRLGSLISPEF	480		
Db	421	PDDLTYLATVLMVCKFGYTAFSWVYYTALYPTVYNNMGVGSSTASRLGSLISPEF	480		
OY	481	VYLGAYDRFLPYILMSGSLITLITAILTLFLESFGTLPDDITDQMLRVKGMKRRKTPSHTR	540		

DB 481 VILGADVRLPYILMGSILITLITLFLPESFGTLPDPTIDQMLRVGKMKHRTPSHTR 540  
QY 541 MKDGOERPTILKSTAF 557  
DB 541 MKDGOERPTILKSTAF 557

RESULT 2  
OCN2\_MOUSE STANDARD: PRT: 557 AA.  
AC 0920E8: 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Organic cation/carnitine transporter 2 (Solute carrier family 22, member 5) (High-affinity sodium-dependent carnitine cotransporter).  
GN SLC22A5 OR OCTN2  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Kidney;  
RX MEDLINE=99113835; PubMed=9916797;  
RA Nezu T., Yamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,  
RA Nikido H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsushita T.,  
RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,  
RA Tsuji A.;  
RT "Primary systemic carnitine deficiency is caused by mutations in a  
RT gene encoding sodium ion-dependent carnitine transporter.";  
RL Nat. Genet. 21:91-94(1999).  
RN [2]  
RP SEQUENCE FROM N.A., AND VARIANT JVS ARG-352.  
RC STRAIN=C3H;  
RX MEDLINE=99057546; PubMed=9837751;  
RA Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;  
RT "A missense mutation of mouse OCTN2, a sodium-dependent carnitine  
RT cotransporter, in the juvenile visceral steatosis mouse.";  
RL Biochem. Biophys. Res. Commun. 252:590-594(1998).  
RN [3]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=99384224; PubMed=10454528;  
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,  
RA Chen J., Conway S.J., Ganapathy V.;  
RT "Functional characteristics and tissue distribution pattern of organic  
RT cation transporter 2 (OCTN2), an organic cation/carnitine  
RT transporter.";  
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).  
CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE  
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE  
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF  
CC CARNITINE.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL  
CC STEATOSIS (JVS).  
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
CC CATION SUBFAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AB015800; BA36590.1; -;  
DR EMBL: AF111425; AAC99787.1; -;  
DR EMBL: AF110417; AAD54060.1; -;  
DR MGD: MGI:1329012; SLC22A5.  
DR InterPro: IPR003662; sub\_transporter.

DR Pfam: PF00083; sugar\_tr.1.  
DR Prosite: PS00216; SUGAR\_TRANSPORT\_1; 1.  
KW Transport; Transmembrane; Glycoprotein; Disease mutation.  
FT TRANSMEM 21 41 POTENTIAL.  
FT TRANSMEM 143 163 POTENTIAL.  
FT TRANSMEM 173 193 POTENTIAL.  
FT TRANSMEM 198 218 POTENTIAL.  
FT TRANSMEM 233 253 POTENTIAL.  
FT TRANSMEM 258 278 POTENTIAL.  
FT TRANSMEM 342 362 POTENTIAL.  
FT TRANSMEM 374 394 POTENTIAL.  
FT TRANSMEM 407 427 POTENTIAL.  
FT TRANSMEM 431 451 POTENTIAL.  
FT TRANSMEM 489 509 POTENTIAL.  
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 352 352 L -> R (IN JVS)  
SQ SEQUENCE 557 AA: 62779 MW: 60933F0EE9612B204 CRC64:  
Query Match 87.2%; Score 2513; DB 1: Length 557;  
Best local similarity 85.5%; Pred. No. 1.8e-153;  
Matches 476; Conservative 39; Mismatches 42; Indels 0; Gaps 0;  
QY 1 MRDYEVTAFLCEKGPORLIFLLSASITNGFTGLSSVLTITPERRCYPPDAALSS 60  
DB 1 MRDYEVTAFLCEKGPORLIFLLSASITNGFTGLSSVLTITPERRCYPPDAALSS 60  
QY 61 AMRNHVPRLRDREVPVSHCRRYRLATIANFSALGELPGRDVDLGOLESCIDGMEFS 120  
DB 61 AMRNHVPRLRDREVPVSHCRRYRLATIANFSALGELPGRDVDLGOLESCIDGMEFS 120  
QY 121 QDYVSTIVTEKNLYCEDDKAPLTISLFFVGYLGSTISQGLSDRGKRVLFVYGMQ 180  
DB 121 KDVFSTIVTEKNDLVCKDKAPLTLSLFFVGYLGSTISQGLSDRGKRVLFVYGMQ 180  
QY 181 TGRSFLQIFSKNFEFVYLVVGKQISNYAFAVLSTELGKSVRIFFSTLGCIFYA 240  
DB 181 TGRSFLQIFSKNFEFVYLVVGKQISNYAFAVLSTELGKSVRIFFSTLGCIFYA 240  
QY 241 FGVMVPLFAFVIRDMRLVALIMPVLCVLAWFIPESPRMLISQGRFEAEVITRKA 300  
DB 241 FGVMVPLFAFVIRDMRLVALIMPVLCVLAWFIPESPRMLISQGRFEAEVITRKA 300  
QY 301 AKANGIVPSTTFDPSELDDLSKKQOSHNLIDLRWNIMVYIMSTIMLTISVGYFG 360  
DB 301 AKANGIVPSTTFDPSELDDLSKKQOSHNLIDLRWNIMVYIMSTIMLTISVGYFG 360  
QY 361 LSLDPNLHGDIIVNCFLSAMVEVPAYVLAWLLQYLPRRYSMAVAFLGSSVLLFMQV 420  
DB 361 LSLDPNLHGDIIVNCFLSAMVEVPAYVLAWLLQYLPRRYSMAVAFLGSSVLLFMQV 420  
QY 421 PPDLYIATLVVWVKFGVTAFAFMYVYTAELYGVVNNQGVSSYASLSGLISPYE 480  
DB 421 PSELFLSTALVWVKFGVTAFAFMYVYTAELYGVVNNQGVSSYASLSGLISPYE 480  
QY 481 VILGADVRLPYILMGSILITLITLFLPESFGTLPDPTIDQMLRVGKMKHRTPSHTR 540  
DB 481 VILGADVRLPYILMGSILITLITLFLPESFGTLPDPTIDQMLRVGKMKHRTPSHTR 540  
QY 541 MKDGOERPTILKSTAF 557  
DB 541 MKDGOERPTILKSTAF 557

RESULT 3  
OCN2\_RAT STANDARD: PRT: 557 AA.  
ID OCN2\_RAT  
AC 070594; 09QML0;  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (rel. 40, Last annotation update)  
DE Organic cation/carnitine transporter 2 (Solute carrier family 22,  
DE member 5) (High-affinity sodium-dependent carnitine cotransporter)  
DE (USP2r) (Cn1).  
GN SLC22A5 OR OCTN2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=9820080; PubMed=9541011;  
RA Schoenig E., Spitzenderger F., Engelhardt M., Martel F., Oeding N.,  
RA Gruendemann D.;  
RT "Molecular cloning and characterization of two novel transport  
RT proteins from rat kidney."  
RL FEBS Lett. 425:79-86(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;  
RX MEDLINE=99011422; PubMed=9792817;  
RA Sekine T., Kishihara H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y.,  
RA Kanai Y., Endou H.;  
RT "Molecular cloning and characterization of high-affinity carnitine  
RT transporter from rat intestine."  
RL Biochem. Biophys. Res. Commun. 251:586-591(1998).  
RN [3]  
RP SEQUENCE FROM N.A.; AND CHARACTERIZATION.  
RX MEDLINE=99384224; PubMed=10454528;  
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Lebach F.H.,  
RA Chen J., Conway S.J., Ganapathy V.;  
RT "Functional characteristics and tissue distribution pattern of organic  
RT cation transporter 2 (OCTN2), an organic cation/carnitine  
RT transporter."  
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).  
CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE  
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE  
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF  
CC CARNITINE.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES  
CC AND IN THE GLOMERULI IN THE KIDNEY, IN THE MYOCARDIUM, VALVES, AND  
CC ARTERIOLES IN THE HEART, IN THE LABYRINTHINE LAYER OF THE  
CC PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE  
CC BRAIN.  
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
CC CATION SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AJ001933; CA05106.1; -;  
DR EMBL: AB017260; BA34399.1; -;  
DR EMBL: AF110416; AA054059.1; -;  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF00083; sugar\_tr.1;  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR TRANSPORT: Transmembrane; Glycoprotein.  
FT TRANSMEM 21 41  
FT TRANSMEM 143 163 POTENTIAL.  
FT TRANSMEM 173 193 POTENTIAL.  
FT TRANSMEM 198 218 POTENTIAL.  
FT TRANSMEM 233 253 POTENTIAL.  
FT TRANSMEM 258 278 POTENTIAL.  
FT TRANSMEM 342 362 POTENTIAL.  
FT TRANSMEM 374 394 POTENTIAL.  
FT TRANSMEM 407 427 POTENTIAL.

FT TRANSMEM 431 451 POTENTIAL.  
FT TRANSMEM 489 509 POTENTIAL.  
FT CARBOHYD 57 57  
FT CARBOHYD 64 64  
FT CARBOHYD 91 91  
FT CONFLICT 275 275  
SQ SEQUENCE 557 AA; 62567 MW; 172472E7B0A5F043 CRC64;  
Query Match 87.0%; Score 2509; DB 1; Length 557;  
Best Local Similarity 85.5%; Pred. No. 3.3e-153;  
Matches 476; Conservative 42; Mismatches 39; Indels 0; Gaps 0;  
QY 1 MDDYEVAFIAGEMGPQRLIFELISASTINGFGLSVFLIATPERHCVPDAANLSS 60  
DB 1 MDDYEVAFIAGEMGPQRLIFELISASTINGFGLSVFLIATPERHCVPDAANLSS 60  
QY 61 AMRNHTVPLRLDGRVPHSCRRYRLATIANFSALGLEPGRVDVIGOLESCDGMERS 120  
DB 61 AMRNHSIPLETGRQVPOSCRRYRLATIANFSALGLEPGRVDVIGOLESCDGMERN 120  
QY 121 QDVYSTIVTEKNLVCEDDKAPLISFPFVGVLGSTRISGOLSDRFGKVLFTMGMO 180  
DB 121 KDVFLLSTVTEWDVCKDKWKAPLITSLFVGVLMSGFISGOLSDRFGKVLFTMGMO 180  
QY 181 TGFSPLOIFSKFEMFVFLVVGQISNYAAFLGTEILGKSVRIJFTLGCIFFYA 240  
DB 181 TGFSPLOIFSVFEMFVFLVVGQISNYAAFLGTEILGKSVRIJFTLGCIFFYA 240  
QY 241 FGYVYLPFAVYFIRDMRLVALIMPVLCVALMFIPESPRMILISGRFEAEYITKA 300  
DB 241 FGFVYLPFAVYFIRDMRLVALIMPVLCVALMFIPESPRMILISGRFEAEYITKA 300  
QY 301 AKANGIVPSTIFDPSELODSSKKOOSHNLIDLRNNIRVATMSINLWMTISVGYG 360  
DB 301 AKFNGIVAPSTIFDPSELODSSKKQSHHITDVRTNIRITIMSLIWLITISVGYG 360  
QY 361 LSLDPNHLGDIYVNCFLISAMVEPAVYLAWLLOYLPRRSMATLFLVGSVLLFMOLY 420  
DB 361 LSLDPNHLGDIYVNCFLIAAVEPAVYLAWLLOHLPRRYSISALFLVGSVLLFIOLY 420  
QY 421 PDDYVLAIVLVWGKFGYTAFAFVWYVYTAFLYVYVNNKGVGSYASRIGSTISYF 480  
DB 421 PSELFLSTALVWVGFGITSAYVYVYTAFLYVYVNNKGVGSYASRIGSTISYF 480  
QY 481 VYLGAVDRLPYILMGSILITLITLILFLPESFGPLPDTIDQMLRVGKMKRKPSTR 540  
DB 481 VYLGAVDRLPYILMGSILITLITLILFLPESFGAPLPDTIDQMLRVGKMKRKPSTR 540  
QY 541 MUKDQGERPTLIKSTAF 557  
DB 541 TOKDGESEPTVLKSTAF 557  
RESULT 4  
YLXS CAEEL STANDARD: PRT; 751 AA.  
AC P46501;  
DT 01-NOV-1995 (rel. 32, Created)  
DT 01-NOV-1995 (rel. 32, Last sequence update)  
DT 01-NOV-1995 (rel. 32, Last annotation update)  
DE Hypothetical 84.8 kDa protein F23F12.5 in chromosome III.  
GN F23F12.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Du Z.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

```
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U12965; AAA20607.1; -
CC WormPep; F23P12.5; CE01252.
CC InterPro; IPR002184; Sfb.
CC Pfam; PF02175; Sfb; 1.
CC Pfam; PF00083; sugar tr; 1.
CC
CC KW Hypothetical protein; Transmembrane.
CC
CC FT TRANSMEM 24 44 POTENTIAL.
CC FT TRANSMEM 57 77 POTENTIAL.
CC FT TRANSMEM 102 122 POTENTIAL.
CC FT TRANSMEM 137 157 POTENTIAL.
CC FT TRANSMEM 200 220 POTENTIAL.
CC FT TRANSMEM 235 255 POTENTIAL.
CC FT TRANSMEM 320 340 POTENTIAL.
CC FT TRANSMEM 348 368 POTENTIAL.
CC FT TRANSMEM 379 399 POTENTIAL.
CC FT TRANSMEM 410 430 POTENTIAL.
CC FT TRANSMEM 432 452 POTENTIAL.
CC FT TRANSMEM 515 535 POTENTIAL.
CC FT TRANSMEM 547 567 POTENTIAL.
CC FT TRANSMEM 583 603 POTENTIAL.
CC FT TRANSMEM 614 634 POTENTIAL.
CC FT TRANSMEM 678 698 POTENTIAL.
CC
CC SO SEQUENCE 751 AA; 84832 MW; A6C4F3540295EFC CRC64;

Query Match 13.4%; Score 387.5; DB 1; Length 751;
Best Local Similarity 25.8%; Pred. No. 1e-17;
Matches 106; Conservative 89; Mismatches 189; Indels 27; Gaps 9;

QY 127 TIVTEMNIVC-EDMKAKPLTSLFVGVLLGSFISGQLSDPFGKKNVLFVGMQGTFSF 185
DB 303 SKVDDFKKFGCTKAYDAWVATFIOFVLGATVYGHLDHFGKPKVPSFGISVGLFGV 362
QY 186 LQIFSKNEMFVFLVGVGMQISNYAAVFLGTETLLKSVRIESTLGVCIFFAFGY-- 243
DB 363 ASGFAPSEVFAAFRIYVGTISILIVFYVILFEIPEDEGRVLR-----FENMGYAR 417
QY 244 MVLPLFAFIRDMRLVALTMPGVLCVAMWFIPESPRMILISQGREFEAEVITRKAKA 303
DB 418 LVETLACFICGYMSAAIATSLPLIPVLLIPESPKMENTKKRPDARAARAVAWL 477
QY 304 NGIYVPSFIIPDSELOD-LSSKKQOOSHNIIDLRTWNRMTVMSIMMWTISVGYGLS 362
DB 478 SGIPVYNDSDOSIEISEKLEKSKRTYTMKDLFTSWTAFTVIVGSISLWSTLSAFGSD 537
QY 363 LDPNMLADIFVNGFLSAMVEVPVYVLAAILLQYLP-----RRYSM--ATAFLGGSYLLF 416
DB 538 LNSGLAENFLISQFVSGAVTAPAKIFVLLDYVPSDRRLRMLQYQIMILCYIMV 597
QY 417 MOLVPP-----DLYYLATVLMWGKESVTAASFMSVYTAELVPTVVRNKGVS 467
DB 598 LMIPESDCSGSGSDLAII--IINIIGVSFIETWDCYLVAAVECPFKIRITIGTCS 655
QY 468 TASRGLSLSPFYVIGAYDRFLPYILMGS-ITLITLITLFLPESGTP 517
DB 656 LIARIGALLAQMAVLSDIYRAPYAVCSIGTISLISCVLPDTRGVDL 706

RESULT 5
ID YMP3 CAEEL STANDARD; PRT; 1222 AA.
AC 010947;
DT 01-OCT-1996 (Rel. 34, Created)
```

```
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 139.9 kDa protein B0361.3 in chromosome III.
GN B0361.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Du Y.;
RL Submitted (FEB-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U00031; AAK1864.1; -
CC WormPep; B0361.3; CE00752.
CC Pfam; PF00083; sugar tr; 1.
CC
CC KW Hypothetical protein; Transmembrane.
CC
CC FT TRANSMEM 51 71 POTENTIAL.
CC FT TRANSMEM 148 168 POTENTIAL.
CC FT TRANSMEM 182 202 POTENTIAL.
CC FT TRANSMEM 232 252 POTENTIAL.
CC FT TRANSMEM 263 283 POTENTIAL.
CC FT TRANSMEM 307 327 POTENTIAL.
CC FT TRANSMEM 329 349 POTENTIAL.
CC FT TRANSMEM 355 375 POTENTIAL.
CC FT TRANSMEM 382 402 POTENTIAL.
CC FT TRANSMEM 413 433 POTENTIAL.
CC FT TRANSMEM 442 462 POTENTIAL.
CC
CC SO SEQUENCE 1222 AA; 139868 MW; CBA42A80A254FB80 CRC64;

Query Match 12.4%; Score 357; DB 1; Length 1222;
Best Local Similarity 21.1%; Pred. No. 1.5e-15;
Matches 118; Conservative 106; Mismatches 209; Indels 126; Gaps 14;

QY 3 DYDEVIAFLGEMKPPQRLIFPLLSASITPNGTGLSSVLLA-----TPHRKRVDPANL 58
DB 35 DPKFEVAYGAYGKYQIFVYVQLT-----NFEYSMSMYTMSFVQINLEKQCE----- 83
QY 59 SSAMRNHTVPLRLRDREVPHSRRYRLATIANFSALGLEPGRDVDLGOLEQESCLDGE 118
DB 84 ----YKNETIP-----ISETQ-----IETESKRAFGMLNNEYC-----G 114
QY 119 FSQDYYL-----STIVTEMNIVCEDMKAKPLTSLFVGVLLGSFISGQLSDPFGK 170
DB 115 IAEINTLVNVTNOKASNTLVDPDLSCHMFEQFGELTIFIGAVVAPPMMLADRYGRK 174
QY 171 NVLFVIMGKGTGSLQIIFSKNEMFVFLVGVGMQISNYAAVFLGTETLLKSVRIITF 230
DB 175 PIIVTATTAIFLANMAASFPNPAIFLLRAFGACSDSYLSVASVATCEYSEKRA-AM 233
QY 231 STGVCCEFYAFGVNVPLEFAFIRDMRLVALTMPGVLCVAMWFIPESPRMILISQGRF 290
DB 234 IIVYVNAVMSLGAVMLLVLTMDDMRKRYFVLSGVGFALMYLTPSPMLITKNKT 293
QY 291 EEAEVITIRAKAKANGIVBSTIFDPSLODLSKKQOOSHNIIDLRTWNRMTVMSIML 350
DB 294 EKLKRYIKTANR----- 305
QY 351 WMTISVGYFGLSDTPNMGDIFVNGFL-SAMVEVPVYVLAAILLQYLPRRYSM----- 403
DB 306 -NVISLVEFAISPMVSVELGD-OVQAFVYSLIEIPAGLAVIPLMKMKRMKIVTWCIVE 363
```

QY 404 ATALFGSVLLFMOLVPPDLVYLVAVGKFGVTAASVYVYTAELPTVAVNNMG 463  
 DB 364 QTLALGVVFL-----DSTEFKIVIMLVAKVNAITIIYSHVPIWATEQFTSVRSICF 416  
 QY 464 GVSSTASRLGSLSPFYVYLGAIDRFLPYILMGSLTILTAIITLPLPSFGTLPDITDQ 523  
 DB 417 SLMNIFQSGIIMSPYVKIIVMSPMWIPVIALFSPISATILAFILHETKNNKKLPTDIES 476  
 QY 524 MLRVKGMKHKRTPSHRML 542  
 DB 477 L-----SYSPSETNDL 486

RESULT 6  
 YQUL\_CAEEL STANDARD; PRT; 529 AA.  
 AC P30638; Q21101;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 58.3 kDa protein ZK637.1 in chromosome III.  
 GN ZK637.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=92168156; PubMed=1538779;  
 RA Sultoni J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,  
 RA Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A.,  
 RA Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,  
 RA Ainscough R., Waterston R.;  
 RT "The C. elegans genome sequencing project: a beginning.";  
 RL Nature 356:37-41(1992).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN-BRISTOL N2;  
 RA Durbin R.;  
 RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z11115; CAAT7460.1; -.  
 DR EMBL: 222175; CAAT7460.1; JOINED.  
 DR EMBL: Z11115; CAAT7460.1; JOINED.  
 DR PIR: S15786; S15786.  
 DR WormPep: ZK637.1; CE06638.  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF00083; sugar\_tr.1.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
 KW Hypothetical protein; Transmembrane; Transport.  
 FT DOMAIN 1 85 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 86 106 1 (POTENTIAL).  
 FT DOMAIN 107 120 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 121 141 2 (POTENTIAL).  
 FT DOMAIN 142 157 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 158 178 3 (POTENTIAL).  
 FT DOMAIN 179 179 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 180 200 4 (POTENTIAL).  
 FT DOMAIN 201 208 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 209 229 5 (POTENTIAL).

FT DOMAIN 230 237 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 238 258 6 (POTENTIAL).  
 FT DOMAIN 259 319 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 320 340 7 (POTENTIAL).  
 FT DOMAIN 341 392 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 373 373 8 (POTENTIAL).  
 FT DOMAIN 394 410 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 411 431 9 (POTENTIAL).  
 FT DOMAIN 432 434 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 435 455 10 (POTENTIAL).  
 FT DOMAIN 456 461 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 462 480 11 (POTENTIAL).  
 FT DOMAIN 481 482 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 483 502 12 (POTENTIAL).  
 FT DOMAIN 503 529 CYTOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 529 AA; 36317 MW; 8D2FF4CB15EC02D CRC64;

Query Match 12.3%; Score 355.5; DB 1; Length 529;  
 Best Local Similarity 26.5%; Pred. No. 7.7e-16;  
 Matches 140; Conservative 97; Mismatches 186; Indels 105; Gaps 24;

QY 52 VPDAANLSSAMRNHTVPLRLDGRVPHSCRRYRLATIANFSAIGLEP--GRVD----- 104  
 DB 10 VLEASNLTEAVVDLTAKOLKEIRHVGDDE-AVRYSNLDRTTELG-EPTQGRSPDSEKTP 67  
 QY 105 -----LGQLEQF-SCLDGEFESQD-----VYLSIT-----VTENLVLCEDDMKAPLT 145  
 DB 68 TVDEAVELGFRPQLKSLSTIGAMAMADMEMMLSLSPALACEMGT---SSVQALV 124  
 QY 146 ISLFFVGLGSLFISGQLSDRFQ-RKNVLFVT-----MGNGTFS--FLQIFSKNFEMP 196  
 DB 125 TTCVFGSMMLSTTWGICDRFGRKGLFTSLVACMGVSGMSPHRYVLLFFRGLTGR 184  
 QY 197 VLFVLVGMQISNYAAV--LGTLLGKSVRIITSTLCVCFYAGVWVLPFAFIR 254  
 DB 185 -----GIGGVPOSVTLVAEFLPTAORACVYLIES-----FWAIGAVEALLVYFVM 231  
 QY 255 D---WRMLVALTMP-GVLVALWMPFESPRLVSGRFEAEVIRKAKANGIVPS 310  
 DB 232 ESFWRKALMFLSLPLGIFAVASFW-LPRESARFMAAGHERALEITQAAARRNRQVLP 290  
 QY 311 TTFPSELQDSSKKQOSHNLDD-LRTWNIRVWTMSIMLWMTISVGFGLSDTPML 368  
 DB 291 G-----RIVSSTKAGSESRGDIANLSPDLRTKRTILLMCIAWTFASYYGWLFTTVL 343  
 QY 369 -----HGDIFVN-----CPLSANVEVPAYVLMILLQYLPRTYSMA 404  
 DB 344 FQSHDECHGGILFSGTQMEVCCPLTRSDYFDLSTTIAEPGLITVILITIEWGRKRTYA 403  
 QY 405 T--ALFLGGSVLLFMOLVPPDLVYLVAVGKFGVTAASVYVYTAELPTVAVNNMG 462  
 DB 404 LEAVFAIFFLFLYFCFD-----RTTYVLLFVVARAFISGAFQCAVYDTPVRYTTIRAVG 459  
 QY 463 GVSSTASRLGSLSPFYVYLGAIDRFLPYILMGSLTILTAIITLPLP 510  
 DB 460 LGTCSAMARIGAIVA-----SEKSLPLPIGIYGTATILGLIASLSLP 501

RESULT 7  
 YGCS\_ECOLI STANDARD; PRT; 445 AA.  
 AC Q46909;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical metabolite transport protein YGCS.  
 GN YGCS OR B2771.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxId=562;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655.
RX MEDLINE-97426617; PubMed=9278503.
RA Blatcher F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.
RA "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U29579; AAA69281.1; ALT_INIT.
DR EMBL; AE000360; AAC75813.1; ALT_INIT.
DR EcoGene; EG13126; Y9CS.
DR InterPro; IPR003662; sub transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
DR Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 23 43 1 (POTENTIAL).
FT DOMAIN 44 56 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 57 77 2 (POTENTIAL).
FT DOMAIN 78 85 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 86 106 3 (POTENTIAL).
FT DOMAIN 107 114 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 115 135 4 (POTENTIAL).
FT DOMAIN 136 142 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 143 163 5 (POTENTIAL).
FT DOMAIN 164 175 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 176 196 6 (POTENTIAL).
FT DOMAIN 197 253 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 254 274 7 (POTENTIAL).
FT DOMAIN 275 286 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 287 307 8 (POTENTIAL).
FT DOMAIN 308 311 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 312 332 9 (POTENTIAL).
FT DOMAIN 333 337 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 338 358 10 (POTENTIAL).
FT DOMAIN 359 369 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 370 390 11 (POTENTIAL).
FT DOMAIN 391 400 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 401 421 12 (POTENTIAL).
FT DOMAIN 422 445 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 445 AA; 48234 MW; B59E452721B15774 CRC64;

Query Match 10.7%; Score 308; DB 1; Length 445;
Best Local Similarity 26.6%; Pred. No. 6.9e-13;
Matches 102; Conservative 79; Mismatches 168; Indels 34; Gaps 11;

```

```

QY 323 SKKQSHNILLRTWNIIMVITMSIMLMTISVGYFGLSDTPNLHGDFVNCFLSAMV 382
DB 231 TATHKHIKTLFSSRYWR---RTAFNSVFVCLVPIPMFVYTWLPITAGIGLEDAFLFSL 287
QY 383 EYPAVYLAAILLOYPDRRYSMATALFLGGSVLE-----MQLVPPDLYLATVYWGKF 437
DB 288 MLNALLIVGALLG-LVLFHLLARKEFLGSPFLATVVMACLPSSGSLTLLFLFST 346
QY 438 GYTAASMYVYTAELYPTVYRMGVGSSTARSGLISPYFV--YLGAYDRPLPYIM 495
DB 347 TITAVSNVLGIDPAESPPTDIRSLGVGFATMSRLGAAVSTGLIPVYLAQMGQVYLLIL 406
QY 496 GSILITAILT-LFDPSEGTPL 517
DB 407 ATVILGVYTWIMAPETKALPL 429

RESULT 8
YT13 CAEEL STANDARD; PRT; 435 AA.
ID YT13 CAEEL
AC Q10917;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 48.6 kDa protein B0252.3 in chromosome II.
GN B0252.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae.
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Du Z., Waterston R.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U23453; AAC46757.1; -.
DR WormPep; B0252.3; CE02419.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 55 75 POTENTIAL.
FT TRANSMEM 384 404 POTENTIAL.
SQ SEQUENCE 435 AA; 48571 MW; 5F6160359FA1840B CRC64;

Query Match 10.1%; Score 291.5; DB 1; Length 435;
Best Local Similarity 21.5%; Pred. No. 7.7e-12;
Matches 90; Conservative 78; Mismatches 168; Indels 83; Gaps 8;

```

```

Db 270 ISASDI-----VEDRDENGSSFKTLRE-----292
Oy 367 NHGDIENVNCFISAVNEVPAVY-LAMLLIQYLPRRYSMATALFLGGSVLTQMOLVPPDY 425
Db 293 -----ITFTTTLFQDRRYLYLFWI-----GILLFEGKRPLE 326
Oy 426 YLATVLYVMGKFGVTAASVMVYVYTAELYPVVRNMGVGSSTASRLGSLISPYFVLYGA 485
Db 327 FCA-----HEFRSSSLHFFSDPHEQJLPFDGRKNCIGFCETLSRFGMLSPYLSHLTA 380
Oy 486 YDRFLPYILMSLTILTAITLFLPESFGTPIPTIIDMLRYKMKHKKTPSHRMKLD 544
Db 381 VHALAPATLTLIAVSGGLTLILPETLNTKLPTI-----AETASRRLDID 428

RESULT 9
YCEL_BACSU STANDARD: PRT: 400 AA.
AC 034691;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical metabolite transport protein ycel.
CN YCEL.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunano M., Tamakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation."
RL SUBCELLULAR LOCATION: to the EMBL/GenBank/DBD databases.
CC -1 SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1 SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB000617; BAA22256.1; -.
DR EMBL: 299105; CAB12089.1; -.
DR Subtilist; BG12773; ycel.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT DOMAIN 1 14
FT TRANSHEM 15 35
FT TRANSHEM 36 49
FT TRANSHEM 50 70
FT TRANSHEM 71 77
FT TRANSHEM 78 98
FT TRANSHEM 99 119
FT TRANSHEM 120 142
FT TRANSHEM 143 163
FT TRANSHEM 164 165
FT TRANSHEM 166 186
FT TRANSHEM 187 217
FT TRANSHEM 218 238
FT TRANSHEM 239 253
FT TRANSHEM 254 274
FT TRANSHEM 275 280
FT TRANSHEM 281 301

```

```

FT DOMAIN 302 304 EXTRACELLULAR (POTENTIAL).
FT TRANSHEM 305 325 10 (POTENTIAL).
FT TRANSHEM 326 343 CYTOPLASMIC (POTENTIAL).
FT TRANSHEM 344 364 11 (POTENTIAL).
FT TRANSHEM 365 370 EXTRACELLULAR (POTENTIAL).
FT TRANSHEM 371 391 12 (POTENTIAL).
FT TRANSHEM 392 400 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 400 AA; 43708 MM; E0AE0CE5DD27395 CRC64;

Query Match 10.0%; Score 287; DB 1; Length 400;
Best local similarity 24.9%; Pred. No. 1,4e-11;
Matches 104; Conservative 76; Mismatches 161; Indels 76; Gaps 18;

Oy 116 GWEF-SQDY-YLSTIV-----TENLVCEQ-DKAPLTLISLFFGVLLGSGLSQSDRG 168
Db 19 GWLFDADVDGILSLFTIALHVEWNLSPEEKWIG-----SVNSIGMAAGFLGLADRIG 74
Oy 169 RKNVLEVTMGQGTGFSFLQIFSKNFEMFVYLVYVNGOISNVVAVFYGTETL-----G 223
Db 75 RKNVFTITLFCFSGISAFVTSLSAFLRLRVIGMGAGELPVASTVSEAVPEKRG 134
Oy 224 KSVRIIFSTLGVCIFFAAGYVLPLEAFYI---RDMRLVALTMPGVCLVLMWFIPE 280
Db 135 RVIVLLES-----FWAVGWLAAALISFYIPSPGQOALLLALPAFVALYLRSTLPS 188
Oy 281 PRHLISGREGAEAVIIRKAKANGIVSTIFDSELDLSKKQOS-HNILDLRITN 339
Db 189 PKY-----ESLSAKRSIMENKSVARQY 213
Oy 340 IRNVITMSIMLMTISVGFGLSLDTPN---LHGDIIVNCF-----LSAVVEVPAYLWL 392
Db 214 IRTYVLMST-VWFCVVFSTYGMFLMLPSVMLKGFMSISFEVLLMTLAQIPGFSAM 272
Oy 393 LLDYLRKRSMATAL-LGGSVLLEFMQLVPPDYLYATVLYVMGKFGVTAASVMVYVTA 451
Db 273 LIEKAGRKWLIVYLLGTGSAFFFG--TADSLTLTLAGVILSEFNL-GANGVLYAYTP 329
Oy 452 ELTPVVRNMGVGSSTASRLGSLSPYFV-YIGAVDRFLPYILMSLTILTAITL 507
Db 330 EQFPTAIRATGSETTAFRIGIFGFLVGLTAA--RHISFVSLFICAILAV 384

RESULT 10
YAAU_ECOLI STANDARD: PRT: 443 AA.
AC P31679; P31578; P75628;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1993 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical metabolite transport protein yaaU.
GN YAAU OR B0045.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region."
RL Nucleic Acids Res. 20:3305-3308(1992).
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.D.,
RA Mau B., Shao Y.;

```





CC EMBL: AE000272; AAC74845.1; -  
DR EMBL: D90821; BAA15573.1; ALT\_INIT.  
DR Ecocore: EG13487; ydjk.  
DR InterPro: IPR003662; sub.transporter.  
DR Pfam: PF00083; sugar\_tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;  
KW Complete proteome.  
FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 26 46 1 (POTENTIAL).  
FT DOMAIN 47 60 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 61 81 2 (POTENTIAL).  
FT DOMAIN 82 90 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 91 111 3 (POTENTIAL).  
FT DOMAIN 112 112 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 113 133 4 (POTENTIAL).  
FT DOMAIN 134 153 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 154 174 5 (POTENTIAL).  
FT DOMAIN 175 181 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 182 202 6 (POTENTIAL).  
FT DOMAIN 203 271 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 272 292 7 (POTENTIAL).  
FT DOMAIN 293 301 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 302 322 8 (POTENTIAL).  
FT DOMAIN 323 329 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 330 350 9 (POTENTIAL).  
FT DOMAIN 351 351 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 352 372 10 (POTENTIAL).  
FT DOMAIN 373 399 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 400 420 11 (POTENTIAL).  
FT DOMAIN 421 441 12 (POTENTIAL).  
FT TRANSMEM 442 459 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 459 459 BC8AB53ECB8BD77 CRC64;  
SQ SEQUENCE 459 AA; 49602 MM; 8C8AB53ECB8BD77 CRC64;

Query Match 9.2%; Score 265; DB 1; Length 459;  
Best Local Similarity 26.2%; Pred. No. 4e-10;  
Matches 101; Conservative 79; Mismatches 153; Indels 52; Gaps 18;  
QY 152 GVLGSGISGQLSDRGKRVLFVTMGQTSFSLQIFSKNEMFVLEVLVGQGISNY 211  
DB 70 GMFLGALVGILGDKTGHRNFIYEALHIASMVGAFSPMDELACRFVGVGLALL 129  
QY 212 VAAPVLTGTEIL-GK-----SVRIIFS---TLGVCITFAFGMVLPLRAYFIRDRMLLYA 262  
DB 130 VTLEFGTEYMGRRNGRWSSRVSTIGWSTPLCSLIAMG--LTPPLIS-AEWMNRVOLL 186  
QY 263 LTPGVLGVAL-WMFIPESPMLISQGRFEAEVLIIRK---AAKANGIVPSTIFPSE 317  
DB 187 PALLSLNTALAMRFPPSPMLBESRGKYGEAEKVMRSIEGVAKRQKPLP-----PVV 241  
QY 318 LQDLSKKQOQSHNIDILRTNINRMVTIMSLMWTISVGFGLSDTP-----NLH 369  
DB 242 IAD-DGKAPQAVPSALLTGLVKRLVILGSCVL-IAMNVQYTLINMLPTIFMTQGINLK 299  
QY 370 GDIFNCLSMVVEVP-AVYLAAILLOVLPBRYGMAATLFLGGSVLLFMQLVPPDLVYLA 428  
DB 300 DSIVLNTN--SMFGAPGEGIFIAMLVMDKIPRK-TWVGGLLLIIVLVG-----IYSLQ 349  
QY 429 T---VLVWGRGKVTAASNV-----YVYTALEYPTVYNNMGVGSSTASRLGSLSPV-- 479  
DB 350 TSMLLITLIGFLITFVYMYCYASAVVPPEIWPTEAKLRSSGLANAVRISGIAAPYAV 409  
QY 480 FVYLGAYDRFLPYILMGSLLTITAI 504  
DB 410 AVLLSSYGVTVGFILLGAVSIIVAI 434

RESULT 12  
STV2\_RAT  
ID SYV2\_RAT STANDARD; PRT; 742 AA.

AC 002563;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Synaptic vesicle protein 2 (SV2).  
GN SV2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
NCBI\_Taxid=10116;  
RX [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.  
RC TISSUE=Brain;  
RA MEDLINE=92390722; PubMed=1519064;  
RT Bajjalieh S.M., Peterson K., Shingal R., Scheller R.H.,  
RT "SV2, a brain synaptic vesicle protein homologous to bacterial  
RT transporters.";  
RL Science 257:1271-1273(1992).  
CC -1- FUNCTION: MAY ACT AS A VESICLE-LOCALIZED NEUROTRANSMITTER  
CC TRANSPORTER OR MAY FUNCTION AS AN ION TRANSPORTER OR CHANNEL.  
CC -1- SUBCELLULAR LOCATION: SYNAPTIC VESICLE.  
CC -1- TISSUE SPECIFICITY: NEURAL AND ENDOCRINE CELLS OF BRAIN AND  
CC SPINAL CORD.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
DR EMBL: I05435; AAA42188.1; -  
DR Pfam: PF00083; sugar\_tr; 1.  
KW Synapse; Nerve; Glycoprotein; Neurotransmitter transport;  
KW Transmembrane.  
FT DOMAIN 1 163 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 182 182 POTENTIAL.  
FT DOMAIN 183 204 LUMENAL (POTENTIAL).  
FT TRANSMEM 205 225 POTENTIAL.  
FT DOMAIN 226 242 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 243 260 POTENTIAL.  
FT DOMAIN 261 262 LUMENAL (POTENTIAL).  
FT TRANSMEM 263 281 POTENTIAL.  
FT DOMAIN 282 294 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 295 322 POTENTIAL.  
FT DOMAIN 323 334 LUMENAL (POTENTIAL).  
FT TRANSMEM 335 355 POTENTIAL.  
FT DOMAIN 356 445 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 446 469 POTENTIAL.  
FT DOMAIN 470 592 LUMENAL (POTENTIAL).  
FT TRANSMEM 593 611 POTENTIAL.  
FT DOMAIN 612 626 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 627 647 POTENTIAL.  
FT DOMAIN 648 649 LUMENAL (POTENTIAL).  
FT TRANSMEM 650 669 POTENTIAL.  
FT DOMAIN 670 694 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 695 711 POTENTIAL.  
FT DOMAIN 712 712 LUMENAL (POTENTIAL).  
FT TRANSMEM 713 731 POTENTIAL.  
FT DOMAIN 732 742 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 498 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 548 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 742 AA; 82705 MM; 565DE7EF2929D5DB CRC64;

Query Match 9.1%; Score 262.5; DB 1; Length 742;  
Best Local Similarity 17.5%; Pred. No. 9.4e-10;  
Matches 115; Conservative 103; Mismatches 173; Indels 267; Gaps 21;  
QY 96 GLEPGRDVDIGQLRQESCLDGMFESQDVYISTVTE-----WNL----- 134

```
Dh 125 GLSDGEGPGGREGAORRRKREELAQQ--YETILRECGHGFWNTLYFVLGLAMADGVE 182
Qy 135 -----VCEDDMKAPLTISLFEYGVLLGSFISGOLSDRGRKRVLFVTGMQ 180
Dh 183 VFVVGFLVPSAEKDMCLSDSKMKGLIYIGMVGAFMLGGLADLRGRGCLLISLTVN 242
Qy 181 TGFSEFLQFESKNFEMFVVLVLMGQISNVAAVFLGTEILGKSVRIIFSTLG----- 234
Dh 243 SVFAFFSSEFVGQYGFTEFLCRLLSCVG-----IGGSIPLVSEYSEFLAQE 287
Qy 235 -----VCFYFANG-----YMLPLF-----AFIFIDRMMLVALTMPEVLC 270
Dh 288 KRGEHLMSLWCMFMWIGYAAAMAMAILPHYGSFOMGSAVQFHSWVVFVLEAFPSVFA 347
Qy 271 VALMWFIPESPRMLISOGFEAEVITIRKA-----AKANGIYVPSITIDPSFLQDLSKKQ 326
Dh 348 IGAULTQEPSPRFLLENKHKDEAMMVLKQVHDTMKRAKG--HPRRVSVYHIKTI-----H 401
Qy 327 QSHNILD-----LRTWNI-----RMVTMSIMLMTISVGY 356
Dh 402 QDEFLIEIOSDTGYWORMGVRLSLGQVWGNFLSCFSPRYRITLMMMGVMTFMSFSY 461
Qy 359 FGLSDPPENL-----368
Dh 462 YGLTWPEPDMIRHLQADVAAARTKVPGERVEHVTENFTLENOIHGQYFNDKFGLRL 521
Qy 369 -----HGDIFFNC-----376
Dh 522 KVSFEDSLFECEYEDVTSNTFFRNCPTINIVFYNTDLEFYKVSRLVNSTFLLNKE 581
Qy 377 -----FLSAMVEVPAYVILAMLLQYLPRLRYSNATALLGGSVL-- 414
Dh 582 GCPIDVGTGEGAYMYFVSFLGLAVLPGNIVSALLMDKIGRLMLA-----GSSVLSG 636
Qy 415 ---LEMLQVPPDXLYLAVLMVNGKRGVY--AAFSMVYVYRAELPYVYRMKGVSSTAS 470
Dh 637 VSCFFLSFGNSESAMIA-LCLCFG--GVSIASWMLDVLVLELVPSPDKRTAEPFALMC 693
Qy 471 RLGSIL--SPFYVILGAVDRFLPYILNGSLTILPAITLFPESGPGPLPDTIDOMLR 526
Dh 694 KLAIVLGISITSPVG-----IKKAPILFASAAALGSLALKLEPTEQVLIQ 742

RESULT 13
YFIG_BACSU STANDARD; PRT; 482 AA.
ID YFIG_BACSU STANDARD; PRT; 482 AA.
AC P54723;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical metabolite transport protein yfig.
GS yfig.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=96262713; PubMed=8704981;
RA Yamamoto H., Uchiyama S., Fajar A.N., Ogasawara N., Sekiguchi J.,
RT "Determination of a 12 kb nucleotide sequence around the 76 degrees
RL Microbiology 142:1417-1421(1996).
RL -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D50543; BAA09111.1; -
DR EMBL: Z99108; CAB12655.1; -
DR Subtilisin, Bg11854; yfig.
DR InterPro: IPR003663; Sugar_transporter.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr.1.
DR PRINTS: PR00171; SUGTRANSPORT.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT DOMAIN 1 29
FT TRANSMEM 30 50
FT DOMAIN 51 59
FT TRANSMEM 60 80
FT TRANSMEM 81 92
FT DOMAIN 93 113
FT TRANSMEM 114 120
FT DOMAIN 121 141
FT TRANSMEM 142 155
FT TRANSMEM 156 176
FT DOMAIN 177 184
FT TRANSMEM 185 205
FT TRANSMEM 206 263
FT TRANSMEM 264 284
FT DOMAIN 285 301
FT TRANSMEM 302 322
FT DOMAIN 323 331
FT TRANSMEM 332 352
FT TRANSMEM 353 373
FT DOMAIN 374 400
FT TRANSMEM 401 421
FT DOMAIN 422 423
FT TRANSMEM 424 444
FT DOMAIN 445 482
SQ SEQUENCE 482 AA; 52756 MW; 90851C4FC48EE01 CRC64;

Query Match 8.9%; Score 256; DB 1; Length 482;
Best local Similarity 23.9%; Pred. No. 1.6e-09;
Matches 105; Conservative 83; Mismatches 167; Indels 84; Gaps 19;

Qy 144 LFLSLFEVGLVLSGFSISGOLSDRGRKN-----VLFV--TWGMQTFGFLQIFSKNEM 195
Dh 64 LVASSLLGAFAMGCGRLSDRGRKTIIVLALFLFATIGC-----TESPNASV 115
Qy 196 FVVLVFLVGM--GOISNYAAFLVGLTEILGKSVRI-----FSLGVCIFYAFGYVLP 248
Dh 116 MIAFRLLGLAVGCASVYPTFLAELSPABRGRIYQNELMIVIGQLLAYFNMIGST 175
Qy 249 FAIFIDMRMLVALTMPGVLCYALW---FIPESRMLISOGREAEVITIRKAKANG 305
Dh 176 MGESANVWRMLVATLPVAV--LWFGMLIVPESPRMLAAGRMGDALRVLRQIRE 229
Qy 306 IYVPSITFPESELQDL-----SSKKQGSNNILDLRTNINIRVNT--MSIMLWMT 353
Dh 230 -----DSQAQDEIKELKHAIEGTAKKAGH--DFOEPITRILFGLGIAIVQOIT 278
Qy 354 --ISVGFGLSL-----DTPNLHGDIFFNCFLSAMVEVPAYVILAMLLQYLPRLP 400
Dh 279 GVNISYVYGEILREAGFQTEALIGNI-----ANGVISVIAVINGIMLGVRRRPM 333
Qy 401 ---YSNATLFLGGSVLLTMQVPPDLYLAVLVWVGFGVYTAFSMYVYTAELPTV 457
Dh 334 IGOIGTMTALLGLISIVLEGPALPYVVLSTILFLAFOOTASTVWMLSEIFPMH 393
Qy 458 VRNMGVGVSS---TASRLGISPYFY-VLGYARFLPYILMGSITLITALLFLPES 512
Dh 394 VRGLGKISTFPCLTANFLIGTFPPLLNHIGMSATFFLVAMNITALL--FVKRYVPE 451
Qy 513 FGTPLPDTIDOMLRVYGMK 531
```

DB 452 KGRSL-EQLEHSFQYGR 469

RESULT 14

YDJE\_ECOLI STANDARD: PRT: 452 AA.

ID YDJE\_ECOLI

AC P38055; P77244;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Hypothetical metabolic transport protein ydJE.

GN YDJE OR B1769.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OX NCBI\_taxid=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RA "The complete genome sequence of Escherichia coli K-12.";

RT Science 277:1453-1474(1997).

RL [2]

RN SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE=89357501; PubMed=9097039;

RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y., Siyandatar S., Tagami Y., Takeeda J., Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.;

RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 28.0-40.1 min region on the linkage map.";

RT DNA Res. 3:363-377(1996).

RL [3]

RN SEQUENCE OF 360-452 FROM N.A.

RC STRAIN-K12;

RX MEDLINE=89357501; PubMed=2670682;

RA Jerstrome P.G., Bezjak D.A., Jennings M.P., Beacham I.R.;

RT "Structure and expression in Escherichia coli K-12 of the L-asparaginase T-encoding ansa gene and its flanking regions.";

RL Gene 76:37-46(1989).

RN [4]

RP IDENTIFICATION.

RX MEDLINE=95075659; PubMed=7984428;

RA Borodovsky M., Rudd K.E., Koonin E.V.;

RT "Intrinsic and extrinsic approaches for detecting genes in a bacterial genome.";

RL Nucleic Acids Res. 22:4756-4767(1994).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (potential).

CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL, AE000272; AAC74839.1; -.

DR EMBL, D90820; BAA15560.1; -.

DR EMBL, D90821; BAA15567.1; -.

DR EMBL, M26934; -; NOT\_ANNOTATED\_CDS.

DR EcoGene: EG12369; ydJE.

DR InterPro: IPR003662; sub-transporter.

DR Pfam: PF00083; sugar\_tr.1.

DR PROSITE: P500216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.

DR PROSITE: P500217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.

DR Hypothetical protein; Transport; Transmembrane; Inner membrane;

KW Complete proteome.

FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 21 43 1 (POTENTIAL).

FT DOMAIN 44 57 PERIPLASMIC (POTENTIAL).

FT TRANSMEM 58 80 2 (POTENTIAL).

FT DOMAIN 81 91 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 92 114 3 (POTENTIAL).

FT DOMAIN 115 117 PERIPLASMIC (POTENTIAL).

FT TRANSMEM 118 140 4 (POTENTIAL).

FT DOMAIN 141 152 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 153 175 5 (POTENTIAL).

FT DOMAIN 176 178 PERIPLASMIC (POTENTIAL).

FT TRANSMEM 179 198 6 (POTENTIAL).

FT DOMAIN 199 265 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 266 288 7 (POTENTIAL).

FT DOMAIN 289 297 PERIPLASMIC (POTENTIAL).

FT TRANSMEM 298 320 8 (POTENTIAL).

FT DOMAIN 321 326 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 327 344 9 (POTENTIAL).

FT DOMAIN 345 353 PERIPLASMIC (POTENTIAL).

FT TRANSMEM 354 376 10 (POTENTIAL).

FT DOMAIN 377 388 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 389 411 11 (POTENTIAL).

FT DOMAIN 412 415 PERIPLASMIC (POTENTIAL).

FT TRANSMEM 416 438 12 (POTENTIAL).

FT DOMAIN 439 452 CYTOPLASMIC (POTENTIAL).

SQ SEQUENCE 452 AA; 50318 MW; 5A596DBA229205C2 CRC64;

Query Match 8.9%; Score 255.5; DB 1; Length 452;

Best Local Similarity 24.0%; Pred. No. 1.6e-09;

Matches 114; Conservative 92; Mismatches 160; Indels 109; Gaps 24;

DB 21 YRGLTANFSAALGPEGRVDLGLQEQSCIDCW-EPSQDQVLTSTVTE-WNLVCEDDMK 141

Y 84 YRLATIANFSAALGPEGRVDLGLQEQSCIDCW-EPSQDQVLTSTVTE-WNLVCEDDMK 141

DB 11 YRGLTANFSAALGPEGRVDLGLQEQSCIDCW-EPSQDQVLTSTVTE-WNLVCEDDMK 141

Y 142 APLTISLFFGVGLGFTISQGLSDRGK-----NVLFTMGKQTGFSLQITSKFENF 196

DB 60 AAFISALMF-GYFISLGTGFTGDFGRRAFRINLIVIA-ATGAAPV-----DMV 111

Y 197 VLPL-----VLVGMQISNVAAVFGTEILGKSVRIIF-STLGVCFYAFG 243

DB 112 WLIFRFLMTGKCALIMVGYASTETETIPATVGRK-----WSALSTFGKNSPMLSAIGV 167

Y 244 WLPLFAFYFIKDWEMLLVALTMPGVLCAVALMFLP-----ESPRMLISQGFEEAVIIR 298

DB 168 VVIAFFS-----WRIMFL-----GGISILLAWFLSGKYFIESPRMLAGKQIAGACQLR 218

Y 299 KA-----AKANGIVPSTIFPDPSELQDSSKKQOSHNLIDLKRWNRIMVYIMSLMIMT 354

DB 219 EVEQIIRERSIRLP-----PLTSYQSNKVKYIKGTFWLLFEGEMLFRLVAITVLIAW 273

Y 355 SVGFGLSLDPTNLHGDIYV-----CFLSAMVEVPAYV---LAWLLQYLPFRYSM 403

DB 274 NISLYITTWIPT-----IFNNSGIDVDKSLIMNAVIMIGAPGIFLAALIDHIFPR--- 326

Y 404 ATALFLGGSVLLFMQVLPDLYIATVAVLVWVKFGVTAFSM-VY-----VYTAELIYPT 456

DB 327 ----LFGSTLLIILVAGIYISIQTEWAILLYLGIMIFLIFYCEASAVYIPELMP 381

Y 457 VVRMGGVSTASRLGSLSPYFV---YLAYDRFLPYIMGSLITLTA-ILTLF 508

DB 382 HURLRGSGFVANAGRIYVETPIGVAALLTHYSITVFWLVGLLQALVLSIF 436

RESULT 15

ID PCAK\_ACICA

AC PCAK\_ACICA STANDARD: PRT: 457 AA.

Q43975;

15-JUL-1999 (rel. 38, Created)  
15-JUL-1999 (rel. 38, last sequence update)  
01-MAR-2002 (rel. 41, last annotation update)  
4-hydroxybenzoate transporter.  
PCAK.  
Acinetobacter calcoaceticus.  
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
Acinetobacter.  
NCBI\_TaxID=471;  
[1]  
SEQUENCE FROM N.A.  
RP MEDLINE=94341565; PubMed=8063101;  
RA Kowalchuk G.A., Hartnett G.B., Benson A., Houghton J.E., Ngai K.-L.,  
Ornston L.N.;  
RT "Contrasting patterns of evolutionary divergence within the  
RL Acinetobacter calcoaceticus pca operon."  
RL Gene 146:23-30(1994).  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
CC (potential).  
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: I05770; AAC37151.1; -  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF00083; sugar tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Transport; Transmembrane; Inner membrane.  
FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 35 55 1 (POTENTIAL).  
FT DOMAIN 56 72 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 73 93 2 (POTENTIAL).  
FT DOMAIN 94 101 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 102 122 3 (POTENTIAL).  
FT DOMAIN 123 128 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 129 149 4 (POTENTIAL).  
FT DOMAIN 150 168 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 169 189 5 (POTENTIAL).  
FT DOMAIN 190 191 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 192 212 6 (POTENTIAL).  
FT DOMAIN 213 274 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 275 295 7 (POTENTIAL).  
FT DOMAIN 296 310 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 311 331 8 (POTENTIAL).  
FT DOMAIN 332 338 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 339 359 9 (POTENTIAL).  
FT DOMAIN 360 364 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 365 385 10 (POTENTIAL).  
FT DOMAIN 400 421 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 401 421 11 (POTENTIAL).  
FT DOMAIN 422 426 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 427 447 12 (POTENTIAL).  
FT DOMAIN 448 457 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 457 AA; 49277 MW; 4F5B5F77361A1567 CRC64;

Query Match 8.88; Score 252.5; DB 1; Length 457;  
Best Local Similarity 24.9%; Pred. No. 2.5e-09;  
Matches 113; Conservative 78; Mismatches 177; Indels 85; Gaps 20;

QY 110 QSCSDGWEFSQDYLYSTIYEW--NLVC-----EDDW---KAP 143  
DB 15 QRSIDPAALINDAPLSIV--QMLIAIVCFIYFVDSIDTAMGFIAPALADKGVDSQ 72  
QY 144 L--TISLFFGVLLGSFISGLSDRGKRVNLFVIMGMOTGFSFIQIFSKNFMFVILV 501

DB 73 LPPVSAALGMIIGALVSGPTADRFGRKIYLSMSMLVEGFTLACAYSPINDSLVIRRF 132  
QY 202 IYGMGOIYNVYAAPVGLGTEILGKSVRIIFSTLGVCIPIYARGVWVLLPFAVFIKDW----- 256  
DB 133 LTGIGLGAAMRNATTLTSEYCPARIR---SLVYCMF--CGYMGNAIGGFISSMLIRPF 187  
QY 257 ---RMLLYALTPGVLCVAMWFIPESPRMLISQGR-FEEAEVIRKAA--KANGIVPS 310  
DB 188 GWSHSLFLGWNAPLIMLIVTFPLPESVREFLIYVGNKTKVRQILSRIPAKVQGV---T 244  
QY 311 TIFDSELODLSSKK-----QOSHNTIDLTNTNINIMVYIMSMIMNTISVGEFL 361  
DB 245 EPHVPEEVEYEGTKGVFGMLFSKIVKGYVLAMVTFMGLWMTYLTLSWPLTMAETGA 304  
QY 362 SLDTPNLHGDIIF-VNCFLSANVEVPAYVLAWLLQYLPRR---YSMAATLFGSVLLE 416  
DB 305 SLERAAPLGLGFQFGVLSAL-----FIGAMDRFPNRIIAGFYLAIGIF---AVIYG 355  
QY 417 MQLVPPDLYIATVLYVMYKRGVYTAASMYVYTAELYPVAVNMGVGVSSTASRIGSL 476  
DB 356 QSLISNPTL--LALFTLCAG-IAVNGAOSMPLVSARFPYPOCRATGVAMMSGIGRGAV- 411  
QY 477 SPYFVYLGAYDRFLPYIIMGSLTITLITLLEFL 509  
DB 412 --FGAMIGA-----VLGNMWSFTMLISMLI 435

Search completed: July 17, 2002, 02:15:03  
Job time: 664 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: July 17, 2002, 02:14:09 ; Search time 102.99 seconds

(Without alignments)  
935,607 Million cell updates/sec

Title: US-09-521-195b-3  
Perfect score: 2883  
Sequence: 1 MRDYEVTAFLEGEWGPFORL.....HTRMKDGQERPTLKSTAF 557

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_19:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rviro:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query	Match	Length	ID	Description
No.						
1	2876	99.8	557	4	Q96EH6	Q96EH6 homo sapien
2	2295.5	79.6	564	11	Q9W1N6	Q9W1N6 mus musculu
3	2227	77.2	551	4	Q9H015	Q9H015 homo sapien
4	2218	76.9	551	4	Q14546	Q14546 homo sapien
5	2157	74.8	553	11	Q9R141	Q9R141 rattus norv
6	2148	74.5	553	11	Q9Z306	Q9Z306 mus musculu
7	885	30.7	548	5	Q9YCA2	Q9YCA2 drosophila
8	881	30.6	548	5	Q01384	Q01384 drosophila
9	816	28.3	567	5	Q9VCA3	Q9VCA3 drosophila
10	811	28.1	567	5	Q95R48	Q95R48 drosophila
11	799.5	27.7	577	4	Q96RU0	Q96RU0 homo sapien
12	777	27.0	568	5	Q9U539	Q9U539 caenorhabdi
13	777	27.0	576	5	O02270	O02270 caenorhabdi
14	761.5	26.4	563	5	Q95TW6	Q95TW6 drosophila
15	753.5	26.1	561	5	Q9V616	Q9V616 drosophila
16	740	25.7	554	4	Q9N0D4	Q9N0D4 homo sapien

17	738.5	25.6	556	4	Q75751	Q75751 homo sapien
18	738	25.6	554	4	O15395	O15395 homo sapien
19	735.5	25.5	540	11	Q91WU2	Q91WU2 mus musculu
20	734	25.5	554	4	O15245	O15245 homo sapien
21	733	25.4	593	11	P70485	P70485 rattus norv
22	728	25.3	557	4	Q96KH7	Q96KH7 homo sapien
23	728	25.3	593	11	P97558	P97558 rattus norv
24	727	25.2	555	4	O15244	O15244 homo sapien
25	724	25.1	555	11	Q9R0W2	Q9R0W2 rattus norv
26	723.5	25.1	554	6	O77504	O77504 oryctolagus
27	718	24.9	556	11	Q63089	Q63089 rattus norv
28	716.5	24.9	674	5	Q9YIK2	Q9YIK2 drosophila
29	714.5	24.8	553	11	O70577	O70577 mus musculu
30	711	24.7	554	6	O02713	O02713 sus scrofa
31	710	24.6	535	11	O63314	O63314 rattus norv
32	709.5	24.6	504	5	Q961J5	Q961J5 drosophila
33	709	24.6	556	11	O08966	O08966 mus musculu
34	704	24.4	556	11	Q9R1Q4	Q9R1Q4 mus musculu
35	703.5	24.4	539	4	Q9H2W5	Q9H2W5 homo sapien
36	703	24.4	551	4	Q9Y226	Q9Y226 homo sapien
37	693	24.0	542	4	Q96TCL	Q96TCL homo sapien
38	690.5	24.0	551	11	Q9W7W5	Q9W7W5 mus musculu
39	690	23.9	548	4	Q9Y694	Q9Y694 homo sapien
40	688.5	23.9	551	11	Q99JF0	Q99JF0 mus musculu
41	681.5	23.6	551	11	O88446	O88446 rattus norv
42	675.5	23.4	553	4	Q96S37	Q96S37 homo sapien
43	675.5	23.4	562	13	O57379	O57379 pseudopleur
44	661.5	22.9	537	11	Q91WU9	Q91WU9 mus musculu
45	660.5	22.9	456	4	O14567	O14567 homo sapien

## ALIGNMENTS

RESULT 1  
ID Q96EH6 PRELIMINARY; PRT; 557 AA.  
AC Q96EH6;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SIMILAR TO SOLOTE CARRIER FAMILY 22 (ORGANIC CATION TRANSPORTER), MEMBER 5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG CARCINOMA;  
RA Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC012325; AHI12325.1; -  
SQ SEQUENCE 557 AA; 62735 MW; 9C3E15059F20E6A4 CRC64;

Query Match 99.8%; Score 2876; DB 4; Length 557;  
Best Local Similarity 99.8%; Pred. No. 3 2e-216;  
Matches 556; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MRDYEVTAFLEGEWGPFORLIFLLSASIPNGFTGLSSVFLATPBRHVRVDAANLSS 60  
Db 1 MRDYEVTAFLEGEWGPFORLIFLLSASIPNGFTGLSSVFLATPBRHVRVDAANLSS 60  
QY 61 AWRNHTVPLRLDGRVPHSCRRYRATIANFSLGLEPGRVDLGLBEESCLDGEFS 120  
Db 61 AWRNHTVPLRLDGRVPHSCRRYRATIANFSLGLEPGRVDLGLBEESCLDGEFS 120  
QY 121 QDVLTSTVTEWNLVCEDDMKAPLTISLDFYGVLLGSFISGQSDRGKRNVLFTVMGQ 180  
Db 121 QDVLTSTVTEWNLVCEDDMKAPLTISLDFYGVLLGSFISGQSDRGKRNVLFTVMGQ 180  
QY 181 TGFSPLOIFSKNPFEMFVVLFLVLVGMGOISNYVAFLVGLTEILGKSVRIIFSTGLVCIEYA 240

```

Db 181 TGFSELOIFSKNEMFVLEVLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCFYA 240
QY 241 FGYWVLPFLFAFVIRDMRMLVALTMPGVLCVALMWFIPESPRLISQGRPEAEVITIKKA 300
Db 241 FGYWVLPFLFAFVIRDMRMLVALTMPGVLCVALMWFIPESPRLISQGRPEAEVITIKKA 300
QY 301 AKANGIVPSTIFDPSELDDSSKKQOSHNIIDLRTWNIRMYTMSIMLMNTISVGYFG 360
Db 301 AKANGIVPSTIFDPSELDDSSKKQOSHNIIDLRTWNIRMYTMSIMLMNTISVGYFG 360
QY 361 LSLDTPNLHGDIFVNCFLSAMVEPAVYVAMLLQYLRPRYSMAATLFLGGSVLLFMOLY 420
Db 361 LSLDTPNLHGDIFVNCFLSAMVEPAVYVAMLLQYLRPRYSMAATLFLGGSVLLFMOLY 420
QY 421 PDLIYLAATVLMVWKGFGVTAFAFSMYVYTAELYPVVRNMGVGVSTASRLGSIISYF 480
Db 421 PDLIYLAATVLMVWKGFGVTAFAFSMYVYTAELYPVVRNMGVGVSTASRLGSIISYF 480
QY 481 VYLGAIDRFLPYILMGSLLTILTAITLFLPESFGTLPDITDMLRVKGMKHKRTPSHTR 540
Db 481 VYLGAIDRFLPYILMGSLLTILTAITLFLPESFGTLPDITDMLRVKGMKHKRTPSHTR 540
QY 541 MLKDGQERPTILKSTAF 557
Db 541 MLKDGQERPTILKSTAF 557

```

```

RESULT 2
Q9WTN6 PRELIMINARY; PRT; 564 AA.
ID Q9WTN6 01-NOV-1999 (TReMBLrel. 12, Created)
AC Q9WTN6 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE OCTN3.
GN SLC22A9 OR OCTN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=20568258; PubMed=11010964;
RA Tamai I., Ohashi R., Nezu J., Sai Y., Kobayashi D., Oku A.,
RT "Molecular and Functional Characterization of Organic Cation/Carnitine
RT Transporter Family in Mice."
RL J. Biol. Chem. 275:40064-40072(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL: AB018436; BA078343.1; -.
DR MGD: MGI:1929481; SLC22A9.
DR InterPro: IPR003662; sub-transporter.
DR Pfam: PF00083; sugar_tr.1.
DR PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 564 AA; 63320 MW; C37FDC6395DAD01 CRC64;

```

```

Query Match 79.6%; Score 2295.5; DB 11; Length 564;
Best Local Similarity 80.6%; Pred. No. 6.1e-171;
Matches 435; Conservative 54; Mismatches 48; Indels 3; Gaps 1;
QY 1 MRDYEVTAFLEMGFPORLIFFLSASTIPNGFTGSSVFLATPEHRCRVDPDANLSS 60
Db 1 MRDYEVTAFLEMGFPORLIFFLSASTIPNGFTGSSVFLATPEHRCRVDPDANLSS 60
QY 61 AMRNHTVPLRLDGRREVPHSCRRYRLATTANFSALGLEPGRVDVLDGQLEQSCLDGWEFS 120
Db 61 AMRNHTVPLRLDGRREVPHSCRRYRLATTANFSALGLEPGRVDVLDGQLEQSCLDGWEFS 120
QY 181 TGFSELOIFSKNEMFVLEVLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCFYA 240
Db 181 TGFSELOIFSKNEMFVLEVLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCFYA 240

```

```

QY 121 QDVTSTVTENMLVCEDDMKAPLTISLFPVVLGSEFISGOLSDRGRKNVLEVTMGQ 180
Db 121 KDIFLSTIVTEMDLVCKDDMKAPLTISLFPVVLGSEFISGOLSDRGRKNVLEVTMGQ 180
QY 181 TGFSELOIFSKNEMFVLEVLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCFYA 240
Db 181 TGFSELOIFSKNEMFVLEVLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCFYA 240
QY 241 FGYWVLPFLFAFVIRDMRMLVALTMPGVLCVALMWFIPESPRLISQGRPEAEVITIKKA 300
Db 241 FGYWVLPFLFAFVIRDMRMLVALTMPGVLCVALMWFIPESPRLISQGRPEAEVITIKKA 300
QY 301 AKANGIVPSTIFDPSE---LDDSSKKQOSHNIIDLRTWNIRMYTMSIMLMNTISG 357
Db 301 AKANGIVPSTIFDPSE---LDDSSKKQOSHNIIDLRTWNIRMYTMSIMLMNTISG 357
QY 358 YFGSLDTPNLHGDIFVNCFLSAMVEPAVYVAMLLQYLRPRYSMAATLFLGGSVLLFM 417
Db 358 YFGSLDTPNLHGDIFVNCFLSAMVEPAVYVAMLLQYLRPRYSMAATLFLGGSVLLFM 417
QY 418 QLVPPDYLAATVLMVWKGFGVTAFAFSMYVYTAELYPVVRNMGVGVSTASRLGSIIS 477
Db 418 QLVPPDYLAATVLMVWKGFGVTAFAFSMYVYTAELYPVVRNMGVGVSTASRLGSIIS 477
QY 421 QLVPSDLHYLSTLVMWKGFGVTAFAFSMYVYTAELYPVVRNMGVGVSTASRLGSIIS 480
Db 421 QLVPSDLHYLSTLVMWKGFGVTAFAFSMYVYTAELYPVVRNMGVGVSTASRLGSIIS 480
QY 478 PFVYVLAGAYDRFLPYILMGSLLTILTAITLFLPESFGTLPDITDMLRVKGMKHKRTPS 537
Db 478 PFVYVLAGAYDRFLPYILMGSLLTILTAITLFLPESFGTLPDITDMLRVKGMKHKRTPS 537
QY 481 PFVYVLAGAYDRFLPYILMGSLLTILTAITLFLPESFGTLPDITDMLRVKGMKHKRTPS 540
Db 481 PFVYVLAGAYDRFLPYILMGSLLTILTAITLFLPESFGTLPDITDMLRVKGMKHKRTPS 540

```

```

RESULT 3
Q9H015 PRELIMINARY; PRT; 551 AA.
ID Q9H015 01-MAR-2001 (TReMBLrel. 16, Created)
AC Q9H015 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE OCTN3.
GN SLC22A9 OR OCTN3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Spritzgenberger F., Gruendemann D., Schoenig E.;
RX Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RA -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL: Y09881; CA71007.1; -.
DR InterPro: IPR003662; sub-transporter.
DR Pfam: PF00083; sugar_tr.1.
KW Transmembrane.
SQ SEQUENCE 551 AA; 62155 MW; C827A99AA78C9443 CRC64;

```

```

Query Match 77.2%; Score 2227; DB 4; Length 551;
Best Local Similarity 76.3%; Pred. No. 1.3e-165;
Matches 425; Conservative 57; Mismatches 69; Indels 6; Gaps 3;
QY 1 MRDYEVTAFLEMGFPORLIFFLSASTIPNGFTGSSVFLATPEHRCRVDPDANLSS 60
Db 1 MRDYEVTAFLEMGFPORLIFFLSASTIPNGFTGSSVFLATPEHRCRVDPDANLSS 60
QY 61 AMRNHTVPLRLDGRREVPHSCRRYRLATTANFSALGLEPGRVDVLDGQLEQSCLDGWEFS 120
Db 61 AMRNHTVPLRLDGRREVPHSCRRYRLATTANFSALGLEPGRVDVLDGQLEQSCLDGWEFS 120
QY 121 QDVTSTVTENMLVCEDDMKAPLTISLFPVVLGSEFISGOLSDRGRKNVLEVTMGQ 180
Db 121 QDVTSTVTENMLVCEDDMKAPLTISLFPVVLGSEFISGOLSDRGRKNVLEVTMGQ 180
QY 181 TGFSELOIFSKNEMFVLEVLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCFYA 240
Db 181 TGFSELOIFSKNEMFVLEVLVGMGOISNYVAAFVLGTEILGKSVRIIFSTLGVCFYA 240

```



```

Db 181 TGFSLQIFSIWEMETVLFVIVMGQISNYVAFLGTEILKSVRIIFSLGCTFEA 240
OY 241 FGWVLPFAFIFIDMRLMLVALTMPGVLCAVLMWFIPESPRLISQGRFEAEVYIRKA 300
Db 241 VGMILLPLFAFIFIDMRLMLVALTMPGVLCAVLMWFIPESPRLISQGRFEAEVYIRKA 300
OY 301 AKANGIVPSTIFDPSELQDSSKQOSHIIIDLFRWNTRMVITMSIMAMTISVGYG 360
Db 301 AKANGIVPSTIFDPSELQDSSKQOSHIIIDLFRWNTRMVITMSIMAMTISVGYG 360
OY 361 LSLDPNPHGDAVFNCFSAWEVPAVYVAMLLQYLPREYSMATALFLGSGVYLFMOY 420
Db 361 LSLDPNPHGDAVFNCFSAWEVPAVYVAMLLQYLPREYSMATALFLGSGVYLFMOY 420
OY 421 PDLTYLATVLMVWGKEGVTAFAFMYVYTAELYPVVRNMGVSSVSTASRLGSLISPYF 480
Db 421 PDLTYLATVLMVWGKEGVTAFAFMYVYTAELYPVVRNMGVSSVSTASRLGSLISPYF 480
OY 481 VYLGAVDRFLPYILMGSLTITLITLFLPESFGTLPDPTIDQMLRYKGMKHKRTPSHR 540
Db 481 VYLGAVDRFLPYILMGSLTITLITLFLPESFGTLPDPTIDQMLRYKGMKHKRTPSHR 540
OY 541 MKDGOERPTIKSTAF 557
Db 541 MKDGOERPTIKSTAF 557
OY 557 DKMETEENPKYL-TTAF 551
Db 557 DKMETEENPKYL-TTAF 551

```

```

RESULT 4
014546 PRELIMINARY; PRT; 551 AA.
AC 014546:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE OCTN1.
GN OCTN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 1:
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=98086199; PubMed=9426230;
RA Tamai I., Yabuuchi H., Nezu J., Sai Y., Oku A., Shimane M., Tsuji A.;
RT "Cloning and characterization of a novel human pH-dependent organic
RT cation transporter, OCTN1."
RL FEBS Lett. 419:107-111(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; AB007448; BAA23356.1; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 551 AA; 62177 MW; F5903421C789F60A CRC64;

```

Query Match 76.9%; Score 2218; DB 4; Length 551;  
 Best Local Similarity 75.9%; Pred. No. 6,6e-165;  
 Matches 423; Conservative 57; Mismatches 71; Indels 6; Gaps 3;

```

OY 1 MRDDEVATALGEMGPORLIFFLLSASTIPNGFTGSSVFILANPEHRCVDPDANSS 60
Db 1 MRDDEVATALGEMGPORLIFFLLSASTIPNGFTGSSVFILANPEHRCVDPDANSS 60
OY 61 AMRNHTVPLRLDGRREVPHSCRRYRLATIANFSALEBGRDVLGQLQESCLDGWERS 120
Db 61 AMRNHTVPLRLDGRREVPHSCRRYRLATIANFSALEBGRDVLGQLQESCLDGWERS 120
OY 121 QDVYLSITVTEWNLVCEBDMKVPPLTSLFVGVLLGSFVSGQLSDRFGRKNVLFATMAVQ 180
Db 121 QDVYLSITVTEWNLVCEBDMKVPPLTSLFVGVLLGSFVSGQLSDRFGRKNVLFATMAVQ 180

```

```

OY 181 TGFSLQIFSKNEMFVYLVIVMGQISNYVAFLGTEILKSVRIIFSLGCTFEA 240
Db 181 TGFSLQIFSKNEMFVYLVIVMGQISNYVAFLGTEILKSVRIIFSLGCTFEA 240
OY 241 FGWVLPFAFIFIDMRLMLVALTMPGVLCAVLMWFIPESPRLISQGRFEAEVYIRKA 300
Db 241 VGMILLPLFAFIFIDMRLMLVALTMPGVLCAVLMWFIPESPRLISQGRFEAEVYIRKA 300
OY 301 AKANGIVPSTIFDPSELQDSSKQOSHIIIDLFRWNTRMVITMSIMAMTISVGYG 360
Db 301 AKANGIVPSTIFDPSELQDSSKQOSHIIIDLFRWNTRMVITMSIMAMTISVGYG 360
OY 361 LSLDPNPHGDAVFNCFSAWEVPAVYVAMLLQYLPREYSMATALFLGSGVYLFMOY 420
Db 361 LSLDPNPHGDAVFNCFSAWEVPAVYVAMLLQYLPREYSMATALFLGSGVYLFMOY 420
OY 421 PDLTYLATVLMVWGKEGVTAFAFMYVYTAELYPVVRNMGVSSVSTASRLGSLISPYF 480
Db 421 PDLTYLATVLMVWGKEGVTAFAFMYVYTAELYPVVRNMGVSSVSTASRLGSLISPYF 480
OY 481 VYLGAVDRFLPYILMGSLTITLITLFLPESFGTLPDPTIDQMLRYKGMKHKRTPSHR 540
Db 481 VYLGAVDRFLPYILMGSLTITLITLFLPESFGTLPDPTIDQMLRYKGMKHKRTPSHR 540
OY 541 MKDGOERPTIKSTAF 557
Db 541 MKDGOERPTIKSTAF 557
OY 557 DKMETEENPKYL-TTAF 551
Db 557 DKMETEENPKYL-TTAF 551

```

```

RESULT 5
09R141 PRELIMINARY; PRT; 553 AA.
AC 09R141:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ORGANIC CATION TRANSPORTER OCTN1.
GN OCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 1:
RP SEQUENCE FROM N.A.
RX MEDLINE=20286310; PubMed=10825452;
RA Wu X., George R.L., Huang W., Wang H., Conway S.J., Leibach F.H.;
RT "Structural and functional characteristics and tissue distribution
RT pattern of OCTN1, an organic cation transporter, cloned from
RT placenta."
RL Biochim. Biophys. Acta 1466:315-327(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; AF169831; AAD46922.1; -.
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 553 AA; 62362 MW; E26C815768A14AD CRC64;

```

Query Match 74.8%; Score 2157; DB 11; Length 553;  
 Best Local Similarity 72.7%; Pred. No. 3,8e-160;  
 Matches 405; Conservative 65; Mismatches 83; Indels 4; Gaps 2;

```

OY 1 MRDDEVATALGEMGPORLIFFLLSASTIPNGFTGSSVFILANPEHRCVDPDANSS 60
Db 1 MRDDEVATALGEMGPORLIFFLLSASTIPNGFTGSSVFILANPEHRCVDPDANSS 60
OY 61 AMRNHTVPLRLDGRREVPHSCRRYRLATIANFSALEBGRDVLGQLQESCLDGWERS 120
Db 61 AMRNHTVPLRLDGRREVPHSCRRYRLATIANFSALEBGRDVLGQLQESCLDGWERS 120

```

```

Query Match= 74.5%; Score 2148; DB 11; length 553;
Best Local Similarity 72.2%; Pred. No. 1,9e-15;
Matches 402; Conservative 67; Mismatches 84; Indels 4; Gaps 2;

OY 1 MRDDEVATAGLGMGCGFORLIFELLASITIPNGFGLSSVFLATPPHRCRDVDAANLSS 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MRDDEVATAGLGMGCGFORLIFELLASITIPNGFGLSSVFLATPPHRCRDVDAANLSS 60
OY 61 AMRHVTPPLALRGRREVPHSCRRYRATATANSALGLEPGDVLGLDEESCLDOWESS 120
    :|||:::| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 61 SMRNHSTPLRTEKGROVPOSCRRYRATATANSALGLEPGDVLGLDEESCLDOWESS 120
OY 121 QDVLSTIVTENNVLVCEDDMKAPLTIISFEVGLLSGISQLSDRGRKNVLFVTMGQ 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 KDFILSTIVTENNVLVCEDDMKAPLTIISFEVGLLSGISQLSDRGRKNVLFVTMGQ 180
OY 181 TGSEFLOIEKNEFEMFVFLVYLGKQISNYAAVLCSTELGKSVIITSLGVCIFFA 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TGSEFVLOIEKNEFEMFVFLVYLGKQISNYAAVLCSTELGKSVIITSLGVCIFFA 240
OY 241 FGWVVLDFPFYFIRDMRMLVALTMGVLCVALMFIPESEPMILISGFEFEAVITRRA 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 IGVNVLELFLFIRDMRMLVALTMGVLCVALMFIPESEPMILISGFEFEAVITRRA 300
OY 301 AKANGIVPESTIPDSELODLSKKQOOSHNLIDLRWNIRNATVMSIKLMTITSVGYG 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 AKMNSIYAPAGIFDPLELOEINSLKQOKIITLDFRTFNITIVVAVMLMTISGYFA 360
OY 361 LSLDPTMLHGDIVNCFVNEFSAMVEPAVYVLAIVLLQYLRFRYSNATAFELGSLVEMQV 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 LSLVNPMLHGDVVLNCFSLGLEVPVYFTAMLLKTLTPRRIINGVLFPMGSGVLLIQV 420
OY 421 PPDLYITATVLVWVGKFGVTAAFSWYVYITAEVPTVVRNKGCVSSTSASRLSPFF 480
    | : : : :| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|
Db 421 PEDNFVPSILVLMVLRFGITSAFSMLYVETALYPTLVRNMAVGITSMASRVGSIAPYE 480
OY 481 VYLAVDPRFLPYILMSLVTITAILTLFLPSEFGPLPDDTDQMLRVKKGKHKHKTPESHR 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 VYLAVYKRLPYILMSLVTILGIIITLPPFSEFGVTLPEINLEQMKYKGRGCK---KST 537
OY 541 MLKDGERPPIILKSTAF 557
    : : : :| :|||
Db 538 VSDRESPEKVL-ITAF 553

RESULT 7
OYVCA2 PRELIMINARY: PRT: 548 AA.
AC OYVCA2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ORCT PROTEIN (GH21655P).
ORCT OR CG6331.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20191006; PubMed=10731132.
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gockayne J.D.,
RA Adamaitides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champé M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abtill J.F., Agdayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

```



```

Db 216 LPSAGVAMQMFSSVGFMTAGFAYFIHDMRWLQIATLPLGLFLCYWIIIPSSARMLKMG 275
OY 289 REEEAVIIRKAKANGIYVPTIID-----PELDDSSKKOOSHNIIDLTWNIR 341
Db 276 RDEAEVILIEKAKKEKKEVPEIYEQLVDEVAEKKKODEMAAOSPAATVDDLTRYPMIR 335
OY 342 MYTISIMIMMTISVYEGSLDTPNLHGDIENCFELSAMVEPVYVLAWLLQYLPYR 401
Db 336 RTLLIFFDWFMVNSGVYIGLSMTNTNNGNOLVNMIGPPEIPIETITLLFTLNRMGRS 395
OY 402 SNAITLFLGGSVLLFMQVLPDLYLATVLYVNGKGVTAASFMYVYVYTAELXPYVNM 461
Db 396 ILGCTMVAAGISLATITVPSDMNWLIVACAMIGLATITSSYGYIYSABQFPYVNV 455
OY 462 GVGSTASRLGSIISPYEVLGAYDRFLPYLMSLITLAILTLPLESFGRPLDPTI 521
Db 456 GIGASSMARVARGIAPLKLGLGELMRPLILGALSITAGLSRLPLETLNRMPEPTI 515
OY 522 DQMLRVKMKHKKRTPSHRMLKDG 545
Db 516 ED-----GENFGKRPAPQETAEG 534

```

## RESULT 9

```

OYVCAC3 PRELIMINARY; PRT; 567 AA.
AC OYVCAC3:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG13610 PROTEIN.
GN CG13610.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
OC Ephydroidea: Drosophilidae: Drosophila.
RN NCBI_TaxID=7227;
RX STRAIN-BERKELEY.
RA MEDLINE-20196006: Pubmed-10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe C.R., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.A., Bouck J., Brokstein P., Brotlier P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cleyer S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dev T., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista A.E., Gary N.S., Gerhart W.M., Glasser K.,
RA Foster K., Gabriellian A.E., Gary N.S., Gerhart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostreli A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Schaefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skusek M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

```

```

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL: AE003747; AAF56270.1; -.
DR FlyBase: FBgn0039176; CG13610.
DR InterPro: IPR003662; sub.transporter.
DR Pfam: PF00083; sugar.tr.1.
KW Transmembrane.
SQ SEQUENCE 567 AA; 63166 MW; 0DB979469395E60 CRC64;

```

Query Match 28.3%; Score 816; DB 5; Length 567;  
 Best Local Similarity 32.2%; Pred. No. 1.6e-55;  
 Matches 183; Conservative 121; Mismatches 212; Indels 52; Gaps 9;

```

OY 4 YDEVYAFLEGWGPORLIFELLSASITIPNGFTGLSSVFLIAPPHRCRYPDANLSSAMR 63
Db 3 YDEAIILHGDGFRYQKTIYFLICTSIPVAFHRLAGVFLAMPDFCALPFENGSSYDLP 62
OY 64 NHTVPLRLRDGRVPHPSCRRYRLATIANFSALGLEBGRVDL-----GQL-----EDGS 112
Db 63 THLMNLSYPNER-----CSY-----VDVYTEELNSISRSSNETY 101
OY 113 CLDGWFEFSDVYSLTIVTEMNLCEDDMKAPLITSLFVGLGSPISGLSDRGRRKV 172
Db 102 C-SYYVDRSKYIYNSAVTEMNLVCGRDPMAATSDSLFLMGLVGLGSLVIGQLSDKGRPT 160
OY 173 LEVTMGQGFSEFLQIFSKNFEMFVLLFVLYGMOISNYVAAPVLOTETILGKSVRIEST 232
Db 161 LFSALVQLVFLGVLGVADEYFTYFARLMGATTSVGLVLYVVAEMEVGDKR-LVAG 219
OY 223 LGVCIFYAFGVVPLFAFIFIDWRMLVALMPGLVCAVLMWFPESRMLISGRREE 292
Db 220 IFVMAFFSVGFMLTAVFAFVHDMRWLQIATLPLGLIEFYWYIIPESARMLLGRKDC 279
OY 293 AEVYIRKAKANGIYV-----STIEDPSE-----LQDSSKKOOSHNIIDLTWN 339
Db 280 AINAMQKARFNKVEISDALSLELDEGENSEKAKKKEDELDEGPPSWDLFCFPN 339
OY 340 IRVNTIMSIMIMMTISVGFSLDTPNLHGDIENCFELSAMVEPVYVLAWLLQYLP 399
Db 340 LRKTLILFLDWLVYSGVYIGLSMTNTNNGNOLVNMIGPPEIPIETITLLFTLNRMGR 399
OY 400 RYSMATLFLGGSVLLFMQVLPDLYLATVLYVNGKGVTAASFMYVYVYTAELXPYV 459
Db 400 RSLTLCGLVMAQLSLATVYIIPKHTLIVACAMIGLATITSSYGYIYSABQFPYVNV 459
OY 460 NMGVGVSTASRLGSIISPYEVLGAYDRFLPYLMSLITLAILTLPLESFGRPLDPTI 519
Db 460 NVALGASVAVARISGMAPFLNFIATIMKRPILILGSLTIVAGLSLLDRETHKKRPLE 519
OY 520 TIDQMLRVKMKHKKRTPSHRMLKDG 547
Db 520 TI-----ADGERFGKRPAPQETAEG 534

```

```

RESULT 10
OY95R48 PRELIMINARY; PRT; 567 AA.
AC OY95R48:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SD08136P.
GN CG13610.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta;

```



RL Biochim. Biophys. Acta 1418:239-244(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Mathews L.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA MEDLINE:99069613; PubMed-9851916;  
RX none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology."  
RL Science 282:2012-2018(1998).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
DR EMBL: AF110415; AAF21932.1; -  
DR EMBL: Z83228; CAC70093.1; -  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF00083; sugar\_tr; 1.  
KW Transmembrane.  
SQ SEQUENCE 568 AA; 63505 MW; C0A3E73851F44056 CRC64;

Query Match 27.0%; Score 777; DB 5; Length 568;  
Best Local Similarity 31.4%; Pred. No. 1.7e-52;  
Matches 165; Conservative 118; Mismatches 225; Indels 18; Gaps 6;  
QY 3 DYDEVTAFLGEMSPFORIRIFLLS-ASIPNGFTGLSSVFLIATPEHRCRVPDANLSSA 61  
DB 8 DDFVLEQVGNIGYQIVFFIICLPISLPSASARNIPVGNPPHCHIEPEKEYLRP 67  
QY 62 WNHNTVPLRLDGRVPHSCRRYRLATIANFSALGLEPGRVDLQOLQOESCLDGMERSQ 121  
DB 68 LINDRQIL-----SCQYNEQIINVERAFTSAP-VDTYSDRISLIVPCQNGMDYDN 116  
QY 122 DYLSITVTENNLVCEDDMKAPLISLFEYGVLLGSFISGOLSDRFGRKNVLFYTMGQT 181  
DB 117 STYLDLSVTEFNLCDOQAMLEISTSEYVGSFISGICLGYADFGRRRSFVLLVLI 176  
QY 182 GFSLQIFSKNEMFVVLVYLVNGOISNYVAFAVLGTETLLSKVRIIFSTLGVCIYAF 241  
DB 177 VCGTASSFAKDIIESFIILREFGTGLAPALPQIPIIFCMEFMGNSGR-IFSGIMSLFPGA 235  
QY 242 GYVNLPLFAFYLRMRMLVALTMPGVLCVAMWFIPESPRMLISGPREAEVIRKAA 301  
DB 236 AMALLGVAMFIRRMROLTFPCNAPFAFYIITYFFLPESPRMSVSGKMAKQKQKIA 295  
QY 302 KANG---IVPSTIFDPSELQDLSKKQ--QSHNILLRLTNIRMTVIMSLMTISV 356  
DB 296 KNGKSNVDVDELVDMSKNHONAAEKEKTRSHNVTDLEKTPDLRKTLITYIIVWMAI 355  
QY 357 GFGLSLDTPMLHGDIFVNCFLSAMVEVPAYVLAWLLQYLPRRSMATALFLGGSVLLF 416  
DB 356 IYNGLTIVNSMLPYDDVWSFIINGAVELPGYFVWMLLQACAGRWTLAATMIVGICGVS 415  
QY 417 MOLVPPDIYATVLYVWKGKGYTAFAFSKYYVTALYTVVIRNMGVYSSTASRLSIL 476  
DB 416 AMFMPDGPMLVASASFIKRGVSGFAVIYIFAGELPFTVVRALGMSMSVAGSGILL 475  
QY 477 SPFYVLYGAYDRPLPYIMGSLTILTAIILFLPESFGPLPDITD 522  
DB 476 APHIVNLGKIYKILPLIMGIMLSAGILTFPELIGAPLPMTIE 521

RESULT 13  
ID 002270 PRELIMINARY; PRT; 576 AA.  
AC 002270;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE F52R12.1A PROTEIN.  
GN F52R12.1A  
OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Peloderae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mathews L.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE:99069613; PubMed-9851916;  
RX none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology."  
RL Science 282:2012-2018(1998).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
DR EMBL: Z83228; CAB05732.1; -  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF00083; sugar\_tr; 1.  
KW Transmembrane.  
SQ SEQUENCE 576 AA; 64493 MW; F494EE94A7EBC0B1 CRC64;

Query Match 27.0%; Score 777; DB 5; Length 576;  
Best Local Similarity 31.4%; Pred. No. 1.8e-52;  
Matches 165; Conservative 118; Mismatches 225; Indels 18; Gaps 6;  
QY 3 DYDEVTAFLGEMSPFORIRIFLLS-ASIPNGFTGLSSVFLIATPEHRCRVPDANLSSA 61  
DB 16 DDFVLEQVGNIGYQIVFFIICLPISLPSASARNIPVGNPPHCHIEPEKEYLRP 75  
QY 62 WNHNTVPLRLDGRVPHSCRRYRLATIANFSALGLEPGRVDLQOLQOESCLDGMERSQ 121  
DB 76 LINDRQIL-----SCQYNEQIINVERAFTSAP-VDTYSDRISLIVPCQNGMDYDN 124  
QY 122 DYLSITVTENNLVCEDDMKAPLISLFEYGVLLGSFISGOLSDRFGRKNVLFYTMGQT 181  
DB 125 STYLDLSVTEFNLCDOQAMLEISTSEYVGSFISGICLGYADFGRRRSFVLLVLI 184  
QY 182 GFSLQIFSKNEMFVVLVYLVNGOISNYVAFAVLGTETLLSKVRIIFSTLGVCIYAF 241  
DB 185 VCGTASSFAKDIIESFIILREFGTGLAPALPQIPIIFCMEFMGNSGR-IFSGIMSLFPGA 243  
QY 242 GYVNLPLFAFYLRMRMLVALTMPGVLCVAMWFIPESPRMLISGPREAEVIRKAA 301  
DB 244 AMALLGVAMFIRRMROLTFPCNAPFAFYIITYFFLPESPRMSVSGKMAKQKQKIA 303  
QY 302 KANG---IVPSTIFDPSELQDLSKKQ--QSHNILLRLTNIRMTVIMSLMTISV 356  
DB 304 KNGKSNVDVDELVDMSKNHONAAEKEKTRSHNVTDLEKTPDLRKTLITYIIVWMAI 353  
QY 357 GFGLSLDTPMLHGDIFVNCFLSAMVEVPAYVLAWLLQYLPRRSMATALFLGGSVLLF 416  
DB 364 IYNGLTIVNSMLPYDDVWSFIINGAVELPGYFVWMLLQACAGRWTLAATMIVGICGVS 423  
QY 417 MOLVPPDIYATVLYVWKGKGYTAFAFSKYYVTALYTVVIRNMGVYSSTASRLSIL 476  
DB 424 AMFMPDGPMLVASASFIKRGVSGFAVIYIFAGELPFTVVRALGMSMSVAGSGILL 483  
QY 477 SPFYVLYGAYDRPLPYIMGSLTILTAIILFLPESFGPLPDITD 522  
DB 484 APHIVNLGKIYKILPLIMGIMLSAGILTFPELIGAPLPMTIE 529

RESULT 14  
ID 095TW6 PRELIMINARY; PRT; 563 AA.  
AC 095TW6;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE GH2794AP.  
GN GH2794AP.  
OS

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y, CN BW SP;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Gartin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuncio J., Paclob J., Paragas V., Park S., Phanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celnik S.,  
 RU Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY058475; AAL13704.1; -  
 SQ SEQUENCE 563 AA; 62138 MW; D3E2CB4096A195CB CRC64;  
 Query Match 26.4%; Score 761.5; DB 5; Length 563;  
 Best Local Similarity 32.2%; Pred. No. 2.8e-51;  
 Matches 184; Conservative 124; Mismatches 201; Indels 51; Gaps 14;  
 Oy 3 DYDEVTAFLGEMGPORLIFELLASIIIPNGFTGLSSVFLIATPBHRCRPDANLSSA- 61  
 Db 14 DFDLIVEIGFGRFORRNLYLCLPFLAANSLSVFTAGSPYRCVPECDKLVDAE 73  
 Oy 62 ---WRNHTVP-LRLDGEVPHSCRRYRLATIANFSAL-GLFGRVDLQO----- 107  
 Db 74 YGANWVSIAPGWSKRGHFTPTCERF---VANGDHLESSSDPWSAMPDQCPAENFT 129  
 Oy 108 LEOESCLDGEFESODVYLS---TITWEMNLVC-EDDWKAPLITLSFVGVLLGSFISGL 163  
 Db 130 TETERC-----NQFYGSSERTIVQOMGLQCEPENLKLAFVGTIHRAGLVGTALSGYL 183  
 Oy 164 SDRGRKNV-LEVTMGMO-TGFSFLOIFSKNFEMFVFLVVGMOISNYAAVFLGTEI 221  
 Db 184 ADRGRKHIFLCTIVFPAALTVGA-QALSDYISFLFALLNAGVSGVPLAFITGEM 241  
 Oy 222 LGSVRIIFSTLGVCIYFAGYMWLPFAFVIRDMRLVALTMPGVLYALMWFITPESP 281  
 Db 242 VGPGRKEM-SIIVLNYFAVGEALIGL-SVFLPDWMRQGLALASPLPICAVYMWIDESV 299  
 Oy 282 RMLISQGRFEAEYITIRAKAKANGIVPSTIFPDSLODSSKKQOSH----- 330  
 Db 300 RMLIARNRREGAGYIIRRAKAVNRDISVELMASFKQOEIDAETGQDDEGGLHVKKD 359  
 Oy 331 ----IIDLRTNIRNVTIMSIMTISVGFGLSLDPNLHGDFVNCFLSAMVEVP 385  
 Db 360 KIMLAIVEVAHSHTLMGRVAILLITAVNAIVYIGLSLNTFSLGKXYLPALVCLVEIP 419  
 Oy 386 AYVLAMLLQYLPFRYSMATALFGLSVLLFMOLVPPDLVYLAIVLVMGKFGVTAAFSM 445  
 Db 420 GYSLAMFLFRFGRVAVLSSGLLCSITCVASSGVITIGAMVIVTLFLVKLGTISSEAV 479  
 Oy 446 VYVTAELPYVANNMGVSVSTRASRLGSLISPYFVLYGAYDRPLPYITLGLTITLAIL 505  
 Db 480 IYPTTAMMPLTVIRSGGVASTFARRGAMLAFFVLASTYDPLPLLTGLTSLVAGLL 539  
 Oy 506 TLFLPESFGTLPDITDML 525  
 Db 540 SLLEPTEFNKRLPDVEEAI 559  
 RESULT 15  
 O9Y6L6  
 ID O9Y6L6 PRELIMINARY; PRT; 561 AA.  
 AC O9Y6L6;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CG4630 PROTEIN.  
 GN CG4630.  
 OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bens P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reiert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sviderskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Wortley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT \*The genome sequence of Drosophila melanogaster.\*;  
 RL Science 287:2185-2195(2000).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 DR EMBL: AE003819; AAF58407.1; -  
 DR FlyBase; FBgn0033809; CG4630.  
 DR InterPro; IPR003662; sub\_transporter.  
 DR Pfam; PF00083; sugar\_tr.1.  
 KW Transmembrane.  
 SQ SEQUENCE 561 AA; 62041 MW; 573643481595788F CRC64;  
 Query Match 26.1%; Score 753.5; DB 5; Length 561;  
 Best Local Similarity 33.2%; Pred. No. 1.2e-50;  
 Matches 184; Conservative 120; Mismatches 200; Indels 51; Gaps 14;  
 Oy 3 DYDEVTAFLGEMGPORLIFELLASIIIPNGFTGLSSVFLIATPBHRCRPDANLSSA- 61  
 Db 14 DFDLIVEIGFGRFORRNLYLCLPFLAANSLSVFTAGSPYRCVPECDKLVDAE 73  
 Oy 62 ---WRNHTVP-LRLDGEVPHSCRRYRLATIANFSAL-GLFGRVDLQO----- 107  
 Db 74 YGANWVSIAPGWSKRGHFTPTCERF---VANGDHLESSSDPWSAMPDQCPAENFT 129  
 Oy 108 LEOESCLDGEFESODVYLS---TITWEMNLVC-EDDWKAPLITLSFVGVLLGSFISGL 163  
 Db 130 TETERC-----NQFYGSSERTIVQOMGLQCEPENLKLAFVGTIHRAGLVGTALSGYL 183  
 Oy 164 SDRGRKNV-LEVTMGMO-TGFSFLOIFSKNFEMFVFLVVGMOISNYAAVFLGTEI 221

```
Db 184 ADYGRKHIEFECIVFMALITGA--QALSDYISFLFALINAVGTSVYPLAFIIGVEM 241
QY 222 LKSVRIIFSTLGVCIFFAFGYMLPLPAYFIRDMRLVALTMPGVLVALMWFIPESP 281
Db 242 VGPRRREM-SSTVLNMFYAVGBALLGL-SVFLPDMRQLALSVPLICVAFWLVPESEV 299
QY 282 RWLISQGRFEAEVYIIRKAAKANGIVPSTIFDPELODLSKKQOSH----- 330
Db 300 RWLLARNRREQAGVLIIRRAAVNRDDISVELMASFKEQELDAETQEDDYEGGLHVKKDD 359
QY 331 -----ILDLLRTWNIAMVTIMSIMLMTISVGYFGLSDTPNLHQDIFVNCFLSAMVEVP 385
Db 360 KIMLAVKEVARSHILMGRYAILLLIMAVNAIVYGLSINATSLGKNKYLNFAIVCYEIP 419
QY 386 AYVLAMLLQYLPRIYSATATLFLGGSVLLFMQLVPPDLYLATVYLVNKGEGVTAAFSM 445
Db 420 GYSLAMFLRRRGRVALSGSILLCSITCVASGFVTLGANMLVYTLFLVCKIGITSSFAV 479
QY 446 VYVYAEIYPVVRMGVGSSTASRLGSIISPYFVYLGAYDRFLPYILMGSLTITAIL 505
Db 480 IYTFAEKMPYVIRSGGVGMSTFARFGAMLAPVPLIASYDPLPLFLFGLSLVAGLL 539
QY 506 TLFPESEFGTLPDPT 520
Db 540 SLLPETFNRLKLPDT 554
```

Search completed: July 17, 2002, 02:14:10  
Job time: 656 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2002, 23:46:19 ; Search time 7127.59 Seconds

(without alignments)  
5375.799 Million cell updates/sec

Title: US-09-521-195b-4

Perfect score: 1831

Sequence: 1 cggagagctcttggtcgcct.....aggagagaactcgaagagaa 1831

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hlg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_hlg\_hum.\*  
31: em\_hlg\_inv.\*  
32: em\_hlg\_other.\*  
33: em\_hgtgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result Query  
No. Score Match Length DB ID Description

RESULT	1	1831	100.0	1831	9	ABO15050	ABO15050 Homo sapi
LOCUS	1	1827.8	99.8	3252	9	AF057164	AF057164 Homo sapi
DEFINITION	2	1824.6	99.7	2811	9	BC012325	BC012325 Homo sapi
ACCESSION	3	1824.6	99.7	2811	9	BC012325	BC012325 Homo sapi
VERSION	4	1360	74.3	3007	10	AF01933	AF01933 Rattus no
KEYWORDS	5	1360	74.3	3007	10	AF10416	AF10416 Rattus no
SOURCE	6	1356.8	74.1	3038	10	AB017260	AB017260 Rattus no
ORGANISM	7	1355.2	74.0	1881	10	AF111425	AF111425 Mus muscu
	8	1352.6	73.9	1888	10	AB015800	AB015800 Mus muscu
	9	1350.4	73.8	2621	10	AF110417	AF110417 Mus muscu
	10	1170	63.9	2297	10	AB018436	AB018436 Mus muscu
	11	1069.6	58.4	2166	9	HSPTMP	Y09881 H.sapiens m
	12	1066.4	58.2	2135	9	AB007448	AB007448 Homo sapi
	13	973.2	53.2	2258	10	AF169831	AF169831 Rattus no
	14	966.2	52.8	2083	10	AB016257	AB016257 Mus muscu
	15	966.2	52.8	2277	10	BC010590	BC010590 Mus muscu
	16	516	28.2	700	6	AX182500	AX182500 Sequence
	17	516	28.2	25871	9	AB016625	AB016625 Homo sapi
	18	514.4	28.1	12257	2	AC011429	AC011429 Homo sapi
	19	512.8	28.0	191397	2	AC023861	AC023861 Homo sapi
	20	363.8	19.9	3081	9	HS181760	L81760 Homo sapien
	21	363.8	19.9	89174	9	AC008599	AC008599 Homo sapi
	22	363.8	19.9	146797	9	AC034220	AC034220 Homo sapi
	23	363.8	19.9	191397	2	AC023861	AC023861 Homo sapi
	24	334.6	18.3	700	6	AX183016	AX183016 Sequence
	25	334.6	18.3	191732	2	AL596182	AL596182 Mus muscu
	26	306.4	16.7	93409	2	AC073294	AC073294 Mus muscu
	27	306.4	16.7	171274	2	AL596444	AL596444 Mus muscu
	28	302.4	16.5	154256	2	AC073295	AC073295 Mus muscu
	29	295.8	16.2	107650	2	AC035148	AC035148 Homo sapi
	30	235.2	12.8	3848	9	HSAC000954	AC000954 Homo sapi
	31	235.2	12.8	3848	9	H0M17DC32	L80199 Homo sapien
	32	221.2	12.1	700	6	AX183015	AX183015 Sequence
	33	217.2	11.9	700	6	AX182459	AX182459 Sequence
	34	217.2	11.9	29312	4	AF015958	AF015958 Homo sapi
	35	216.4	11.8	2732	4	DMORCT2	Y12400 D.melanogal
	36	215.2	11.8	2064	3	AY058437	AY058437 Drosophil
	37	215.2	11.8	2564	3	AC020006	AC020006 Drosophil
	38	215.2	11.8	156905	2	AC008204	AC008204 Drosophil
	39	215.2	11.8	199016	3	AE003747	AE003747 Drosophil
	40	215.2	11.8	204442	3	AF268892	AF268892 Homo sapi
	41	204.2	11.2	1734	9	AF210455	AF210455 Homo sapi
	42	201.6	11.0	1864	9	AF210455	AF210455 Homo sapi
	43	201.4	11.0	1870	9	HS077086	U77086 Human organ
	44	200.4	10.9	1901	6	E31258	E31258 Liver-speci
	45	200.4	10.9	1901	6	E31259	E31259 Liver-speci

## ALIGNMENTS

RESULT	1	1831 bp	mrna	linear	PRI 06-FEB-1999
LOCUS	ABO15050	1831 bp	mrna	linear	PRI 06-FEB-1999
DEFINITION	ABO15050	Homo sapiens mRNA for OCTN2, complete cds.			
ACCESSION	ABO15050.1	GI:3242597			
VERSION	ABO15050.1	GI:3242597			
KEYWORDS	OCTN2.				
SOURCE	Homo sapiens kidney cDNA to mRNA.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 1831)				
REFERENCE	Nezu, J.				
AUTHORS	Submitted (28-MAY-1998) Jun-ichi Nezu, Chugai Research Institute				
JOURNAL	for Molecular Medicine, Inc., Gene Search Program, 153-2 Nagai,				
	Nihari, Ibaraki 300-4101, Japan [E-mail:nezu@climed.com,				
	Tel:81-298-30-6211, Fax:81-298-30-6270]				
REFERENCE	2 (sister)				
AUTHORS	Tamai, I., Ohashi, R., Nezu, J., Yabuchi, H., Oku, A., Shimane, M.,				
	Sai, Y. and Tsuji, A.				
TITLE	Molecular and functional identification of sodium ion-dependent,				
	high affinity human carnitine transporter OCTN2				



QY	1681	accatgtgcacgaatcctctaaagagtcataaggaatgaaacacacagaaaaatcccaagtcacaca	1740
Db	1681	ACCATTTGACCAATCTCTAAGAGTCAAGAAAGAAACACAGAAAAATCTCAAGTCAACACA	1740
QY	1741	aggatgtttaaagatcgtgtcaagaagaagcccaacatcccttaaaagcagccttcaaat	1800
Db	1741	AGGAGTTRAAAGATGCTCAGAAAAGCCCAACATCTTTAAAGCACAGACCTTTACAT	1800
QY	1801	cgcctccagtaagggagaactgaagagaa	1831
Db	1801	CGCTTCAGTAGGAGAAACTGAAGAGAA	1831
RESULT 2			
AF057164		3252 bp	mRNA linear PRI 01-JUL-1998
LOCUS	AF057164		
DEFINITION	Homo sapiens organic cation transporter OCTN2 (OCTN2) mRNA,		
ACCESSION	AF057164		
VERSION	AF057164.1		
KEYWORDS	GI:3273740		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 3252)		
JOURNAL	Wu, X., Prasad, P. D., Leibach, F. H. and Ganapathy, V.		
REFERENCE	cDNA sequence, transport function, and genomic organization of		
AUTHORS	human OCTN2, a new member of the organic cation transporter family		
TITLE	Biochem. Biophys. Res. Commun. 246 (3), 589-595 (1998)		
JOURNAL	98289574		
REFERENCE	2 (bases 1 to 3252)		
AUTHORS	Wu, X., Prasad, P. D., Leibach, F. H. and Ganapathy, V.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-APR-1998) Biochemistry and Molecular Biology, Medical		
FEATURES	College of Georgia, 1120 15th Street, Augusta, GA 30912-2100, USA		
SOURCE	Location/Qualifiers		
	1. 3252		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="5"		
	/map="5q31"		
	/cell_line="JAR"		
	/cell_type="choriocarcinoma"		
	1. 3252		
	/gene="OCTN2"		
	222. 1895		
	/gene="OCTN2"		
	/codon_start=1		
	/product="organic cation transporter OCTN2"		
	/protein_id="AAC24828.1"		
	/db_xref="GI:3273741"		
	/translation="MRDYDEVTAFLGNGPEPRFLIFELLSAIIPINGTGLSVELLA		
	TPRECRVDPANLSSAMRNHTVPLRLDGRVPHSCGRYRLATIANSAIGLEPGRE		
	VDLQLEQESCLDGMERSQDYLSTIVTEMLVCEDDMKATLITSEFVGLGSFTIS		
	GQLDRGRKNVLEFVTGMOTGSEFLQFESNPFMYLFLVAMGQISNVAVAFME		
	TEILGYSRIIEFTLGYCITFAFGYRMULPLAAYTRRMRLALVIMRGVCAIMWF		
	IPSPRMILISGRFEBAEYIIRKAKANGIVPSTITPDSLDLSKQDSNALMID		
	LRTNNIRVNTTMSIMLWMTISVGFGLSDTPNLHGLIFVNCFLSAWEVDAVYIAM		
	LQYLPRYSMAATLFLGGSVLEFLMQLVDPDLVYLAIVLVAGKFGYAAFSNVYVT		
	AELPTVVRNMGVSVSTASRLGSLISPYEGLVADSEFLPYIIMGSLITITLITLFE		
	LPSEGFPLPTIOTMLRYVKMKRHKRPSTHRLMKDDGERITILKSTAF"		
BASE COUNT	722 a	835 c	820 g 875 t
ORIGIN			
Query Match	99.8%;	Score 1827.8;	DB 9; Length 3252;
Best Local Similarity	99.9%;	Pred. No. 0;	
Matches 1829;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
1	cggaagctcttggtgcgcctcgtcctcgtcgtcgtcgtcggcggcggtgccccgcg	60	

Db	99	CGAGCGGCTTGGGTCGCCCTGCTGCTGGCTTGGCTGCTGCTGGCGGGCGGGTGCCTCCCGCGC	158
OY	61	caagcgaaagcccgccgcltcccaagaccccaagcccgccgcltlygsccltgaaggc	120
Db	159	CACCGGAAAGCCCGCGCGCTTCCCGACCCCGAGCGCGCTGTGGGCTCTAGAGGC	218
OY	121	ggaatgcgggaacthaagacaaagatgaacgccttctbgyggaatlyggagccctcaagc	180
Db	219	GGCATGGGGGACTAGAGAGAGTGAACCGCTTCTGTGGGAGTGGGGGCGCTTCCAGGCG	278
OY	181	ctcaactctctctgctcagcagccagcatatcccaatgctcaltcaacgacctgtcc	240
Db	279	CTCATCTTCTCTCGTCAAGCGAGCATCCCAATGGCTTCAACGGCTGCTGCTCC	338
OY	241	gtgttctcgaatagagaccgccggagacgcgtgcgcgggtgcggaagccgcaactgac	300
Db	339	GTGTCTCTGANTAGGACCCCGGAGACCGCTGCGGGGTCCGAGCGCGCAACTGTAGC	398
OY	301	agcgccctgycgaaccccaactgtcccaactgcgctcgcggaagccgcgagatgcccac	360
Db	399	AGCGCTGGGGSCAACCACTGTGCCACTGTGGGAGGCGCGAGGTGCCAC	458
OY	361	agctgcgcgcgtacacgcgtgcgcacacatgcacaactctcggcgcctcggagctgaacg	420
Db	459	AGCTGCGCGCGCTACCGGCTTCCGACACTGSCCAACTTCTCGGCGCTCGGCTGAGCGG	518
OY	421	ggggcgcaactgagacccctggggcgagcctgaggaagaagatgctcgaatgagcgaactc	480
Db	519	GGGCGCGACCTGGACCTGGGGGACGCTGAGACAGAGAGCTGTGGATGGCTGGGAGTTC	578
OY	481	agtaagaagcttctaactcgttccaccatttgaaccagatgaaacctggtgtgaagaagac	540
Db	579	AGTCAGGACGCTACTGTCACCACTTGAGACAGATGGAACCTGGTGTGATGAGAGCAC	638
OY	541	tgggaagccccactcaaaactccttcttcttcctggtgtgtgtcgttggcctcctaatt	600
Db	639	TGGAGGGCCCACTCACAACTCTCTTGTCTTGTGGGTGCTGTGGGCTCTTCAATT	698
OY	601	tcaaggaagcctgtlcaagaagaagtttggccggaagaactgtctgttccgtgaacatggacgt	660
Db	699	TCAAGGCGGCTGTACAGACAGGTTTGGCGGGAAGAAATGTCTTGTGTACATGGGCAATG	758
OY	661	cagacagcctcagctctctctgcagactcttcctcgaagaatttgagatgttgtcgtcgt	720
Db	759	CAGACAGGCTTACACTTCTTCGAGAACTTCTCGAATAATTTGGAATGTGTCTGCTGTG	818
OY	721	tttctccttgaagcatggggcagacatctccaactatgtagagagcattgtccttggagaca	780
Db	819	TTTGTCTTGTAGGCAATGGGCGAGATCTCAACTAATGTGACACATTTGTCTTGGGACAA	878
OY	781	gaaatccttggcaagtcagttcgtataaataatctcactagctttagagatgtagataattat	840
Db	879	GAATCTTGGCAAGTCAGTTCGTAATAATTTCTACCTTGGAGGTGACATATTATAT	938
OY	841	gcatttgcataagatggcgccacagtttgcattactatccagagactggagagatgtgt	900
Db	939	GCAATTTGGCTACATAGTGTGTCGCACTGTTCCTGCAATAATTCGAGACTGGGATGTGTG	998
OY	901	ctgtgtgcctcgtgaagatagccgggggtgtcgtgtgcgtgcactgtgtgtgtltaacccctag	960
Db	999	CTGTGGGGCGTGAAGATGCCGGGGGTGTGTGTGGCACTGTGGGTGTATCTCTGTAG	1058
OY	961	tccecccgatggtcatalcttccaaggaagatttgaagaagcagagtgatcaatccgcaag	1020
Db	1059	TCCCCCGATGGCTCATCTTCAAGGAGCACTTTGAAGAGCAGAGGTGTATCTCCGCAAG	1118
OY	1021	gctgcacaagaacatggagatctgtgtgccttccaactatctttagccgagtgagttacaa	1080
Db	1119	GCTGCGCAAGCAATGGGATTTGTGTGTGCTTCCACTATCTTTTAGCCCGAGTATGTACAA	1178
OY	1081	gacctaagttccaagaagcagcagctccaacaattctgactctgtcttccaactcgtgat	1140
Db	1179	GACCTAAGTTTCCAAGAGCAAGAGTCCCAACAATCTTGATGTGCTTGCAGACCTGGAT	1238

Qy	1141	atggagatggtgcacatcatatgcatatgtcgttgatgacatatgctggcgcatcttt	1200
Db	1239	ATCCGGATGGGTACCATCATATGTCATAATAGCTGTGGATGACCATTATCATGTGGCTATATTT	1298
Qy	1201	ggccttcgccttgacacacacctactacttgatctggagacatctctgtgaactgcctcttca	1260
Db	1299	GGGGCTTTGGCTTGATACATCCCTAACCTTGGCATGGGAGCATCTTTGTGAACGTGGTCTTTTCA	1358
Qy	1261	ggcagatggtttgaagtcccaagcatatgtgttgcgcctgcctgcctgcctgcataattgcgccg	1320
Db	1359	GGGATGGTTGAAGTCCACAGCATATGTGGCTTGCGCTGGCTGCATATTTTGGCCCGG	1418
Qy	1321	cgctatccatgycacatgcccctctctccctggatggcaatgtccctctctcatcttcacgtg	1380
Db	1419	CGCTATTCCATGGCCACATGCCCCCTTCCGGGTGGCAGTGTCTCTCTCATATGACGTG	1478
Qy	1381	gtaccaccaagacttgttatatttggcctacacagtcctcgtgtgatgtagggcaagtttgaagtc	1440
Db	1479	GTACCCCCACAGACTTTATATTTATTTGGCTTACAGTCCGTGGTATGGTGGGCAAGTTTGGAGTGC	1538
Qy	1441	acggctgcctcttccatgctgaatgctgtagtacaagaagcagcgttatccacaatgggtgaga	1500
Db	1539	ACGGCTGCCCTTTTCCATGTGTACTACTGTATACACAGCCGAGCTGTATCCACAGTGGTGA	1598
Qy	1501	aaacatgggtctgtggagagtcagctccacagaatccgccttggcagacatcctgtctccctac	1560
Db	1599	AACATGGGTGTGGAGTGCAGCTCCACACATCCGCCCTGGCAGCATCTGTCTCTCTAC	1658
Qy	1561	ttcgtttacacctgtgagctacacacgccttccctgcgcctcaatctctatggaaagtctgac	1620
Db	1659	TTTCGTTTACCTTGGTGTCTACGACCGCTTCTGCCCCCTACATTTCTATGGAAAGTGTGAC	1718
Qy	1621	atcctgacaagccatccctactccttgttctccagaagaagcttcgtatccccaatcccaagac	1680
Db	1719	ATCCGATGACAGCCATCTCTACACTTGTTCCTCCAGAGAGCTTGGTATACCCCAATCCCAAGAC	1778
Qy	1681	accatctggccagatgtctgaagttcaagaagaatgaagaacacagaanaacctccaagttaacaa	1740
Db	1779	ACCATTTGACCGATCTTAAGATCTAAAGGAAATGAATACACAGAAAACTCCAAAGTCAACAA	1838
Qy	1741	aggatgtttaaagaatggtctcaagaagaagcccaaatctcttaaagaagacagccttctaact	1800
Db	1839	AGGATGTTAAAAAGATGGTCAAGAAAGGCCACAAATCTTTAAAAAGCAGACGCTTTAAACAT	1898
Qy	1801	cgctcccaagtaaggagaaacttgaagaagaa	1831
Db	1899	CGCTTCAGTAAGGAGAAACTGAAGAGGAA	1929

RESULT	3
LOCUS	BC012325
DEFINITION	BC012325 Homo sapiens, Similar to solute carrier family 22 (organic cation transporter), member 5, clone MGC:20139 IMAGE:4548603, mRNA, complete cds.
ACCESSION	BC012325
VERSION	BC012325.1 GI:15147377
KEYWORDS	MGC.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 2811)
TITLE	Strausberg,R.
JOURNAL	Direct Submission Submitted (09-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCMD/DRP/Gazdar  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susanna Chan, Readman Chiu, Chris Felli, Erin Garland, Ran Guin,  
 Leticia Hsiao, Martin Kirzysinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott,  
 Michael Thorne, Miranada Tsai, Natasja Van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/URL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 29 Row: a Column: 13  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3273740.

FEATURES	Location/Qualifiers
source	1. .2811

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:20139 IMAGE:4548603"
/tissue_type="Lung, large cell carcinoma"
/clone_11b="NH_MGC_18"
/lab_host="DH10B-R"
/notes="Vector: pOT87"
172.1845

```

```

/product="Similar to solute carrier family 22 (organic
cation transporter), member 5"
/protein_id="AAH12325.1"
/db_xref="gi:15147378"
/translation="MRDYDEVTAFLGEMGPORILFILLASIIPINGEFTGLSVFLIA
TPRRCVYPDPAASASMRNTVPLRLDGDGEVHSQRRLATIANFSALGEEIGSD
VDLQLEDESCPDGMEFSQDYILSTIYTENNLVCEDMKAKPLITSLFPGVLLGSFTG
GQLSDREGKRNKLVFTVMGMQGFESFLDIENSNEMFVLLVLMGQGISNVAVFVLTG
TEILGKSVRIIEFTLGVCIFAFYGNVLTPLAYIRDMRLMLATIMDGVCALVLMF
IPESPRMLISQGRFEPAEVIIRKAAKNGLTVPSTIDPSLQDLSLKKQOOSHIIIDL
LRTNNIRKATIMSLMLMTISVCFGLSLDPLNLHGLIPNVCFLISAAVEVPAYLANL
LDQLLPRTSMATALFLGVSLLFPMQLVPPDLYLANLVAVGKGVTAESMYVYVY
AELPVTVRNMGVGSSTASRLGISLSPYEVGAYRFLPYILMGSLTITLTALTLPF
LPSEFGNLPPTIDQMLRVKMKHRRKTPSHRLMKDQERTILTKSTAF"

```

Query Match 99.78; Score 1824.6; DB 9; Length 2811.

```
Best Local Similarity 99.88; Pred: No. 0;
Matches 1827; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

1 cggacggtccttggtcgccctgctgacctgacctgctcgcgcggtgcccgcgcg 60

49 CGGACGGTCTTGGGTCCGCTGCTGGCTTGGCTGGGTGGGCGGCG 108

61 caccqcaaaqccccqccqcttcccaqaccccaqccqccctctgtgqccctctgaqgc 120

109 CACGCGCAAGCCCCGCCGCTTCCCCGACCCAGGCCGCGCTCTGTGGCCCTTGAGGGC 168

121 qacatacqaqactacqaacqaatgaaccaccttctgaqcaqatgaqqqcccttccaqac 180

169 GGCATGGGGACTTAGGACGAGTGGACCGCTTCCTGGGGAGTGGGGGCGCTTCCAGCGC 228

181 ctcatctctcctacacacacacatcaccacaatacctatcctcc 240

[illegible]

211 ++++++ 200

[illegible]

QY 301 agcgctggcgcaacacacactgtccacttgaggctggcgagcgccgaggtgtcccccac 360  
 |||  
 Db 349 AGCCCTGGCGCAACCACTGTCCACTCGGCTGGGAGACGCCGAGAGTGGCCAC 408  
 QY 361 agctgcgcgcgtacacgcctcgcacacatcgccaactctcggcgctcggcggtgagcg 420  
 |||  
 Db 409 AGCTGGCGGCGCTACGGGCTGGCCACCATCGCAACTTTCGGGGCTTGGGCTGGAGCCG 468  
 QY 421 gggcgagcgttgagccttgaggcagcttgagcagagagagcgtctgtgagtgagtgatc 480  
 |||  
 Db 469 GGGCGCACGCTGAGCTGGGGCACCTGGAGCAGGAGAGCTGTCCGGATGGCTGGGATTC 528  
 QY 481 agtaagagcgttaccctgtccaccattgtgacagagtgagagcctggtgtgtgagagagac 540  
 |||  
 Db 529 AGTAAAGACGCTACCTGTCCACCATGTGTGACCGAGTGGAACCTGGGTGTGTGAGAGACG 588  
 QY 541 tggagagcccccacacacatctcctgttctcgttgagggtgtgtgtgtgtgtgtgtgtgt 600  
 |||  
 Db 589 TGGAAAGGCCCACTCACAAATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 648  
 QY 601 tcaaggcagcgttgagcagaggtttggcccgagagaaatgtgtgtgtgtgtgtgtgtgtgtgt 660  
 |||  
 Db 649 TCAGGGCAGCTGTGAGACAGGTGTGGCCGGGAAGATGTGCTTTCGTGACCATGGGATG 708  
 QY 661 cagagcagcgttcaagcttctcagatcttctcagagaaatttgagatgtgtgtgtgtgtgtgt 720  
 |||  
 Db 709 CAGACAGGCTTACGCTTCTGCAAGATCTTCTGAAAGATTTTGAGATTTTGTGTGTGTGT 768  
 QY 721 ttgttctcttgtagcagtgaggcagagatctcacaactatgttgagcagatctgttcttgagg 780  
 |||  
 Db 769 TTTGTGCTTGTAGGAGTGGGCGAGATCTCCAAATATGTGGCAGATTTGTCTGGGAGCA 828  
 QY 781 gaaattcttgcaagtcagctgtcgtataataatctcactgaagtaagagtgatataatttat 840  
 |||  
 Db 829 GAAATTTCTGGCAAGTCTAGTTCGATATATATCTTCTACGTAGGAGTGTGATATTTAT 888  
 QY 841 gcaattgtctacatgtgtgtcgaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 900  
 |||  
 Db 889 GCATTGTGGCTACATGT 948  
 QY 901 ctgtgtgcgttgagcagatgcgcgggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 960  
 |||  
 Db 949 CTGTGTGGGCTGACGATCCGGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1008  
 QY 961 tccccccgatgtcatctctcagagagagatttgagagagagagagagagagagagagagag 1020  
 |||  
 Db 1009 TCCCCCCGATGTGCTATCTCTAGGAGAGATTGAGAGAGAGAGAGAGATTCGCCAAG 1068  
 QY 1021 gctgcgaagcgaatggagatgt 1080  
 |||  
 Db 1069 GCTGCCAAGGCCAATGGGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1128  
 QY 1081 gacctaagtctcagaagaagcagagtcaccaacacatctgagatctgtcgcgaactggaat 1140  
 |||  
 Db 1129 GACCTAAAGTTCACAAGAGCAGCAGTCCCAACAATCTGTGATCTCTTGCAGACCTGGAA 1188  
 QY 1141 atccggatgtgtacacataatgtcacaataatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1200  
 |||  
 Db 1189 ATCCGGATGTGACCATGTCATATGTCATATGTCGTGATGACCATATACATGGGACTATTT 1248  
 QY 1201 gggcttgccttgatatctactacttgcaatggagagacatctgtgaaactgtctcttca 1260  
 |||  
 Db 1249 GGGCTTTCGCTTGATACCTCTAATTTGCATGTGGGACATTTTGTGAACGTCTTCCCTTCA 1308  
 QY 1261 gegaatgttgaagtccacagacataatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320  
 |||  
 Db 1309 GCGATGTGTGAAGTCCACATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1368  
 QY 1321 cgtatctcatggcgaactccctctcctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1380  
 |||  
 Db 1369 CGCTATTTCATGCGCACTGCCCTCTTCTGTGGGGGAGAGTCTTCTTCTCATGACAGCTG 1428  
 QY 1381 gtacccccagactgtatattatgtgctaacagtcctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440

Db 1429 GTACCCCCAGACTTGATATTTGGCTACAGTCTGTGATGTGTGGCAGATTTGGAATC 1488  
 QY 1441 agcgctgcttcttcatggtctacgttataacagcgcgcgcgtgtatcccaaggtgtgaga 1500  
 |||  
 Db 1489 AGGCGCTCTTTTCCATGATGTCTACGTGTACGTGTACAGCGGAGCTGTATCCCAAGTGTGAGA 1548  
 QY 1501 aacatggtgtgtgagtcagctcccaacagatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1560  
 |||  
 Db 1549 AACATGGGTGTGGAGTCAAGCTCCACAGCATCCCGCTGGGAGCATCTGTCTCCCTAC 1608  
 QY 1561 ttgcttacctgt 1620  
 |||  
 Db 1609 TTGCTTTACTTGT 1668  
 QY 1621 atcctgtacagcactcctcactgt 1680  
 |||  
 Db 1669 ATCTGACAGCCATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1728  
 QY 1681 accattgacacagatgtctaaagatcaagaatgaacacagaaaaactccaaagtacaca 1740  
 |||  
 Db 1729 ACCATTGACCAAGTGTCTAGAGTCAAGAGATGAACACAGAAAACTCAAGTCAACACA 1788  
 QY 1741 aggatgtlaaagaatgtcaagaagaaggcccaaatccttaaaagcagccttctaact 1800  
 |||  
 Db 1789 AGGATGTAAAGATGTCAAGAAAGGCCACCAATCTTAAAGCACAGCCTTCAACAT 1848  
 QY 1801 cgtctcagtaagaggaagaactgaagaaga 1831  
 |||  
 Db 1849 CGCTTCCAGTAAAGGAGAACTGAAGAGAA 1879

## RESULT 4

RNAJ1933 3007 bp mRNA linear ROD 30-MAR-1998  
 LOCUS Rattus norvegicus mRNA for putative integral membrane transport protein (USTR2r).  
 DEFINITION

ACCESSION AJ001933  
 VERSION AJ001933.1 GI:3004486  
 KEYWORDS integral membrane transport protein; USTR2r gene.  
 SOURCE Norway rat.

ORGANISM Rattus norvegicus  
 Eukaryota; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 3007)

REFERENCE 1. Gruendemann, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-OCT-1997) D. Gruendemann, University of Heidelberg, Department of Pharmacology, Im Neuenheimer Feld 366, 69120 Heidelberg, FRG

REFERENCE 2. (bases 1 to 3007)  
 AUTHORS Schomig, E., Splizenberger, F., Engelhardt, M., Martel, F., Ordling, N. and Gruendemann, D.  
 TITLE Molecular cloning and characterization of two novel transport proteins from rat kidney

JOURNAL FEBS Lett. 425 (1), 79-86 (1998)  
 MEDLINE 98200080  
 FEATURES

source Location/Qualifiers  
 1..3007

/organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /tissue\_type="kidney"  
 87..1760

gene /gene="USTR2r"  
 87..1760  
 /codon\_start=1  
 /product="putative integral membrane transport protein"  
 /protein\_id="CA05106.1"  
 /db\_xref="GI:3004487"  
 /db\_xref="SPTREMBL:O70594"  
 /translation="MRDYDEVTAFLGEGPFRQLIFLLASLIIPNGFMGMSIVFLAG

CDS  
 1381..1440

TPERCLAPHTVNLSAMRNSTIPIELKDGQVQSPRRYALITNAFSAIQLGPRGD  
 VDLQELQENCDGWENKDYELSTIYTBENDJQDKMKAPLITTSLEFVGLKMSFIS  
 GOLSDRGGRKRIYFLTMGMQGESEFLOLESYNEEMFVLEVLVGMGQISNVAFAVLG  
 TELLSKTSRIITFAVLGCIITAFGEFMLPLAFTIRWRMLLALIVPGVIGCALWME  
 IPESPRMLISQIGRKEAEVILIRKAKNGVIAVSTIPDSELDLUNKRPOSHIYDL  
 VRTNRIINITITMILITLVSVGFSLETPNPHMGIIYVNCFLAAVEPAVYALML  
 LLOHLPRYSISALFLGGSVLLEIQLESFLEITLALVWGEFGTSPASAMYYAT  
 AELPVPVVRNMGVGSSTASLGLSILPEYVIGAYDRFPLPYILMGLTITLTITLITLF  
 FPSEFGAPLPTIDOMIRVKRGIKQMOISQOTRQKDGSESPVLKSTAF"

Query Match	74.3%	Score 1360;	DB 10;	Length 3007;
Best Local Similarity	85.1%;	Pred. No. 1.4e-237;		
Matches 1519; Conservative	0;	Mismatches 265;	Indels 0;	Gaps 0;

QY	48	ggggcccgccgagcagcgcaagccgcccgcgtctccagaccgccgagccgcgtcgtg	107
Db	11	ggccctcagcggcccccgtctccgcgttcggggtatcccccaatccccgagggggtttgtg	70
QY	108	ggacctgagggcgagcatgcgagactacgaagaggtacacgctctcctgtggagattggg	167
Db	71	ggagcctgagagacggcatcgcgagctacgagcaggtacccgcttcttggggagttgggg	130
QY	168	ggccttcagagcgctcatctcttctcgtctcagcgccagcatcatcccaatgcttcaac	227
Db	131	ggcccttcagacgcctcattttcttcttcgctacgagccagcatcattcccccaatggcttcaa	190
QY	228	cgagcgtcgtccgcgtgttcctcgtatagacaccccgagacacgctgcccgggtgcsagaagc	287
Db	191	tggatgtgtctatcgttggttcctgaggggagaccgccgagacacggttgcttggctctcaac	250
QY	288	cgcgaaacctgagcagcgccctgagcgaaacacaactgtcccactgcyggtgcyggaagcgcg	347
Db	251	cgtgaacttgagacacacggctggcccaacacacagattccccgttgagagcaagagacgagcg	310
QY	348	cgagagtgccccaagctctccgcccgtacacgggtccgcacacatccgaactcttcgcgct	407
Db	311	acaggtgctctcagacgttcccccctctacacactgagccacacatccgcaacttctcgcgct	370
QY	408	cgaggctggaagccggggcgcgagctgtagaccttgaggagcagctggaagcagagctgtctgga	467
Db	371	ggggcctggagccggagcggagcgtgagacgttgagacgttgagacagaggaactgccttga	430
QY	468	tggcttgggagattcagtcagtagagctcactcgtccaccattgtgacgagctggaacctggt	527
Db	431	tggcttggggatnacaacaagagacgcttctcgtccacacatcgtgacagagtgaggacttgg	490
QY	528	gtgtgagagcagctgtaagagcccaactcacaatctctgtctcttcgtgggtggtcgtc	587
Db	491	gggtgaagatgacttggaagagcccccacactcacacactcttctgttttgcgggggtgctgat	550
QY	588	ggagctctcatcttcaagggcagctgtlcaagaaggttgccggaagaaatgtcgtctcgt	647
Db	551	gggcctccttcaatttcgggagcagctctcagacaggtttggctccagaaatgtgctgtttt	610
QY	648	ggacatggagcatggaagacagattcaagctctctgcagactctctcgaagaattttgat	707
Db	611	gacacatgggacatgacagatggcttccacttccgacagctcttcttggaacttgcagat	670
QY	708	gatttgcgtgagctgtctctcgttaagcatgagggcagatctcacaactatgtggagaatt	767
Db	671	gatttgcagagcgtttttgttcctgttggaagtgaggttcagatctccaaactacgtggagcatt	730
QY	768	tgcttcggggagcagaatctctggaagtcagttcgtataaatctctacatgattagagt	827
Db	731	tgtctctgggaaacaataattctttccaaagctcaattgcaattatattccgacacttagag	790
QY	828	ggtcattattatgcatcttggtcagaatgtgtgcgcagcagtttgcttaactcatccgaga	887
Db	791	ttgcataattttatgcattttgcttcaatggctgacctctgttggaatacttcaacagag	850

QY	888	ctggcggaatgcgtcgtcgtggtggtcgtgaagatgcccgggggtgtcgtgtggtcactcgtgtg	947
Db	851	ctggaggaatgctgcgtgcttggccctacgtgacagaggggtgctgtgtggtgcctctgtgtg	910
QY	948	gttcaacccctgagtcctcccccgaatggtcctatctcccaaggaaagatttgaagaagcaagtt	100
Db	911	gtttcatccctgagtcgtcccccctatggtctatgtcttccaagcccgagtttaaaagcagaagtt	970
QY	1008	gatlcatcgcgaagctgtccaaagccaatggatgtgtgtgtccttcaactcctcttgaacc	106
Db	971	gattcatccgcgaagctgtccaaatttcaatgaggtattgttgccgttccactatctgtgaccc	103
QY	1068	gagtgatgtacaagaagccctaattccagaagaagcagatcccaacaactctgtgaacctgtct	112
Db	1031	cagtagtattacaaagcctttaaatttccaaagAACCCtCACgtccaccacatttatgtatgtgtgt	109
QY	1128	tgaaaccttgaaatacggatggtgcacacatcatgttcaataatgtctgtgtgtagtaccatc	118
Db	1091	ccgaacacgaataatcacaggaatcacaccatcatttccattatgtccctgtggtgacccaatc	115
QY	1188	agtggtgactatttgggtcttctgcattgaactccttaacttcagtcaggggaactcttgtga	124
Db	1151	actgggctacttgcgaccttcttcttgacactccttaacttcagtcaggggaactcttgtga	121
QY	1248	ctgtctctcttccagcgataggtgtgtagtaccagaatatgtgtgtgcctgtgcgtctgtga	130
Db	1211	ctgtctctcttctgcagctgtgtgaagAACCCcttatgtttgttgccctgtgcctgtgtgtga	127
QY	1308	atattgtcccggtgactatccatgagccactgcctcttccgtgtggtgaggtgctctct	136
Db	1271	gcacctgccccgggggtratttctatctgtctgtcccttcttctgtgggtgacagtgtcttct	133
QY	1368	cttcaatgactgtgtacccccagaactgtatatttggcttacaagtcctgtgtgtgtgtgtg	142
Db	1331	cttcatatagctgtgtgtgcttgcagaaattgtttactgttccacatgcccctgtgtgtgtgtg	139
QY	1428	caagtttgagtagtcggctgtccttcttccatggtctcaogtctaaagccggaagctgtacc	148
Db	1391	caagttttggaatcacctgtgtccttatgttccatggtcttatgtgtatACCGccgacctgtatccc	145
QY	1488	cacagtgtgtgaaacaatgggtgtgtgagtgacgtccacaagaatcccgctctgtgagacat	154
Db	1451	cactgtgtgtgaaAACATtggtgtgtgggggtgtcacttccacagcatcccgcttggcagcat	151
QY	1548	ccgtgtctcctactgtgttaccttgggtcctaaagacggcttccgtcgtccctacatctcat	160
Db	1511	cctgtctccctacttgttgttaccttgggtcctaaagacggcttccgtcgtccctacatctcat	157
QY	1608	gggaaggtctgaacatcctgtgacagacatcctcaacttgcttctccagaagactgtgtac	167
Db	1571	gggaaggtctgacacatcctgtgacacatcctcaacttgcttctccagaagactgtgtac	163
QY	1668	cccactccagaaacccaattacagagatgtcaagagtcgaaggaatgaaacaacaagaanaac	172
Db	1631	cccattccccgagacacattatccacagatgctcagaggtcCAaAGaAtTAaACAAATGcCAAT	169
QY	1728	tccaagtccacaagaagtgttaaagaatgtgtcaagaagaagcccaaatctcctaagaagac	178
Db	1691	CCAaAGCCAAaCAAGGcCGCAAAaAATGtGTGAGAAACCCCAAGGtGtCTTAAGAGcAC	175
QY	1788	agccttctaactcgtcttccagtaagaagggaaactgaaagagaa	183
Db	1751	AGccttctaactcgtcttccagtaagaaggtgAAaAATGtAAAGGAA	174

RESULT	5
AF110416	
LOCUS	3037 bp mRNA linear ROD 10-SEP-1999
DEFINITION	Rattus norvegicus organic cation/carnitine transporter (OCTN2)
ACCESSION	AF110416 complete cds.
VERSION	AF110416.1 GI:5852403
KEYWORDS	.

	SOURCE	Norway rat.
	ORGANISM	Rattus norvegicus
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	REFERENCE	1. (bases 1 to 3037)
	AUTHORS	Wu,X., Huang,W., Prasad,P.D., Seth,P., Rajan,D.P., Leibach,F.H., Chen,J., Conway,S.J. and Ganapathy,V. Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic Cation/Carnitine transporter
	JOURNAL	J Pharmacol. Exp. Ther. 290 (3), 1482-1492 (1999)
	MEDLINE	99384224
	PUBMED	10454528
	REFERENCE	2 (bases 1 to 3037)
	AUTHORS	Wu,X., Huang,W., Prasad,P.D., Rajan,D.P., Leibach,F.H., Chen,J., Conway,S.J. and Ganapathy,V. Direct Submission Submitted (01-DEC-1998) Department of Biochemistry and Molecular Biology, Medical College of Georgia, 1120 15th Street, Augusta, GA 30912-2100, USA
	TITLE	Location/Qualifiers
	JOURNAL	1..3037
	FEATURES	/organism="Rattus norvegicus"
	SOURCE	/db_xref="taxon:10116"
	gene	1..3037
	CDS	/gene="OCTN2" 101..1774 /gene="OCTN2" /codon_start=1 /product="organic cation/carnitine transporter" /protein_id="AAD54059.1" /db_xref="gi:5852404" /translation="MRDYDEVTAFLGEMGPORLIEFLLSAIIIPNGNSIYELAG TPEHRCPLHTVNLSSAMRNISILETDGRQVOSCRYLATLATANSALLEGRR VDLEQLREBNCLDMEYKRDFLSTIYTEMDLPCKDKAPLTITSFPVGYLMSFIS GQLSDREGKAVLELTMGMQGFSEFLDFSNFEKFVLFLVVGMLQSLNSVAEVLID TEILSKSRIRLIFATLGVCIFYAFGMFWPLPAXYLRDMRLLLTVYGVLGALLMM IPESPRWLISQGVKEAEVILRKAKFNGIYAESTIPDSELQDLNSKKPOSHIYDI VRTNRIRITITSIIMLTIISVGEFGLSDPNLHGDIYNCFLLAAVEVPAYVLAMI LLOHLPRYSISALFLGGSVLFIOLVPSLFYSTLAWVKGITSAVSMTVVYV ALELYPYVRNNQGVCSSTRASKGLSTSPYLYLAGADRFUYLIIMGSLTILAILTLLE FPESHGAPLDPTIIDMLRVKGIKMQIOSRTQRKGSESFTLVKSTAF"
	BASE COUNT	701 a 791 c 750 g 795 t
	ORIGIN	
	Query Match	74.3%; Score 1360; DB 10; Length 3037;
	Best Local Similarity	85.1%; Pred. No. 1.4e-237;
	Matches 1519; Conservative	0; Mismatches 265; Indels 0; Gaps 0.
Oy	48	ggtgcgccggcgagcaacgcgaacccgcgcgttcctccagaccacccaggcgcgcgtctgtg 107
Dd	25	ggccctcagacgcccgcgtttccgccttgcggtatccccagatccccggacgggtgctgtg 84
Oy	108	ggcctctgaggcgcatcgcggaactaacagagtgtaacgcgccttccttgcgagtggag 167
Dd	85	ggagccttagagacgcgatccgggacctagcacacagatgacccgcttcttgccgagtg 144
Oy	168	ggccttcagcgacctatcttcttcctgctcagcgccaagcatcatcccaatggttcaa 227
Dd	145	ggccctccagcgcctcattcttcttcctgcctcagccgacacatcattcccaatggttcaa 204
Oy	228	cggcctgctctccgigtgttctctatgacgaccccgagagccgcgtgcgcggtgcggagc 287
Dd	205	tgggtatgcttatcggtgttccttcggggggacccccggagcacccgttgcttgtccttac 264
Oy	288	cggaaacctgagcagcgcttgcgacaacacacactgcccaactgcgctgcggagcgcg 347
Dd	265	cttggtaacttgacacagcgctgcgcacacacactatcccttggaacacgaagacggagc 324
Oy	348	cgaagtgcaccacaagctgcgcgcgcctacgcgctgcgcacacatgcgaacttcgcgcgt 407

Db	325	ACAGTGTCCCTCAAGAGTGGCGGCGGCTACCGAGACTGGCCACCATGCGCAACTCTCTGCGGT	384
Qy	408	cgagctgagacgggggcgcgcacgtggaacctgaggacgtggaagcagaagaagcgtgctgtga	467
Db	385	GGGGCTGAGACCCGGGACGGGAGCGTAGACTTGAGAGCGTGGAGCAGGACGAGAAAGAACTGCTTGA	444
Qy	468	tgagctggaaattcaatcagaagaagcttacctgtccaccattgtgacccgaatggaaacctggt	527
Db	445	TGGCTGGAGATCAACAAAGAGCGTCTCTGTCACCACTCGTGCACAAGTGGAGACCTGGT	504
Qy	528	gtggaagagacgactgtgaaagcccccactcaactctccctgtctctgtgtgtgtgtgt	587
Db	505	GTTAAGAGATGACTGGAAGAACCCCACTACCAACCTCTTGTGTTTTCTGGGTGTGCTGAT	564
Qy	588	ggagctccttaattccagggcagcgtctcagacaaggtttlbgccggagaagaatgtgtgtctgt	647
Db	565	GGGCTCTTATTGTCGGGAGAGCTGTCAGACAGGTTTGTGTGCAAGAATGNGCTGTTTTT	624
Qy	648	gaaccaatggacatgcagacagagcttcaagcttcctgcagactcttcgaaagaatttgaat	707
Db	625	GACCATTGGGATCAACACTGGCTTCAGGCTTCTCGACGCTCTTCTGTGAACTTCAAGAT	684
Qy	708	gtttgtgcgtctgtttgttcctgttgaggacatgggccaagatctcacaactatlgacaatt	767
Db	685	GTTTACAGTCTCTTTTGTCTCTGTTGGCAATGGGTGAGATCTCCAACTACGGAGGATTT	744
Qy	768	ttctcctgggacagaattctcttgacaagcaattcgtaataattctctacgttaagagt	827
Db	745	TGTCCTTGGGAACAGAAATCTTTCCAAAGTCATTGTGATTTATTCGCACACTTAGAGT	804
Qy	828	gtgcataatttaatgatctgtgctacatggtgtctgcacctgtgtcttaactaatccgaga	887
Db	805	TTGCATTTTATTCAGCAATTGGCTTCATGGTGTGCTGCTGTGTTCCATCACTTCATCAGAG	864
Qy	888	cttgcggaatctctgtgtgagcgtctgcagctgcgcggagggtgtctgtgcgtlbgcaactctg	947
Db	865	CTGGAAGATGCTGTGCTTGTGCTCTACCTTACCAAGGGTGTGTTGTGGGCTCTCTGTG	924
Qy	948	gttaacccctgagacccccagatgagtcatactctctaaggacgaatttgaagaggcaaggt	1007
Db	925	GTTTCATCCCTGAGTCCCCCGAGTGGCTCATCTCTCAAGGCGGAGTTAAAGGCGAGAGT	984
Qy	1008	gataatccgaagctgtgcacaagccaaatggagttgtgtgcctccactactttgaacc	1067
Db	985	GATCATTCGGCAAAAGCTGCCAAATTCATATGGATGTGTGGCGCTTCCACTATCTTCGATCC	1044
Qy	1068	gaatgagttcaagaactaagtctcagaagaagacgaagccacaagaactctcgaatcgtc	1127
Db	1045	CAGTAGATTTCAAACACTTAAATATTCCAAGAAAGCTCAGTCGACCACTTATATGATCGGT	1104
Qy	1128	tcgaacctggaatcccgagatgtgtaacatactgtccaaatagatgctgtgaatgacacatc	1187
Db	1105	CCGAACACAGAAATTCAGAGTATCATACCATATGTCATTAATTCCTGTGGTGCACCATATC	1164
Qy	1188	agtgagctatttgggcttgcgtgtgataactcctaactgaatggaggacalcitgtgaa	1247
Db	1165	AGTGGGACTTTCGACTTCTCTTGACACTCTTAATTCGATGGGACATCATATGTGAA	1224
Qy	1248	ctgtctccttcaagcagatgtgtgaatgccagaatatgtgtggcgtgtgcgtgcgcga	1307
Db	1225	CTGCTTCTACTGACACTTGTGAATGCCCGCCGATATGTGTGGCTGGCGTGTGTGATCA	1284
Qy	1308	atatttgcggcgagctattccatcgaatggcactgcctcttcctcggatggacatgtccctc	1367
Db	1285	GCACCTGCCCGGGGATATCTATCTCTGCTGCTCTTCTCTCGGATGGCAATGTCTTCT	1344
Qy	1368	cttcaatcagcttgttaccaccaagactgtatattatgtgctacagtcctcgtgatatgtg	1427
Db	1345	CTTCATACAGCTGTGCTTCAAGAAATGTATTATTTGATTCACATGTCGCCCTGAGATGTGG	1404
Qy	1428	caagtttggagtgacggcgtccttcccatggtctcaagtgtcaagtgtgacacagccgaagctgtatcc	1487
Db	1405	CAAGTTTGGATCACCTCTCCATATTCATGTGTATGTGATGTGACAGCGCGAGCTGTACCC	1464



OY	1488	caacgltgltgaagaacatgggtgtgtgggagtcagctccacagcatccgcgcgttggcagcat	1547
Db	1465	CACGTGTGTGGAAGAACATGGGTGTGGGGGTGACACTTCACAGGATCCCGCTTTGGCAGCAT	1524
OY	1548	ccgtlctccctaacttgcgttttaaccttgggtgcttcagacgcgccttcctgcgcctacatctcat	1607
Db	1525	CCTGTCTCCCTACTTGTGTCTACCTGGGTGCTATGATTCGCTTCCTGCTTATATCTCAT	1584
OY	1608	gggaagtctgaacatctctatgacacacatccctcaactgtttcccaagagagcttcggttac	1667
Db	1585	GGGAGTCTGACCACTCTTGACAGCTATCCCTACACTTGTGTCTTCCACAGAGGCTTTGGGTGC	1644
OY	1668	cccaatcccaagaccccatctgacacagatgtctaagatgcaagaaatgcaacagaaanaac	1727
Db	1645	CCCTCTCCCCACACCCATTATACAGATGCTCAGAGGTCAAGAGATATAAACAATGGCAAT	1704
OY	1728	tccaagtccaacaagagatgttaaaagatgttcaagaagaagccacaatctcttaagaagc	1787
Db	1705	CCAAAGCCAAACAAAGGACGCCAAAATAATGTGTGAGAGAAACCCAAAGGTCCTTAAAGGCAC	1764
OY	1788	agcctcttaacatgcgtcttcagtaagaaggaacatgaagagaa	1831
Db	1765	AGCCTTCTAACACCCCTGTCTCCAGAAAGGTGAAGAACTGAAGGAAA	1808

RESULT	6
LOCUS	AB017260
DEFINITION	Rattus norvegicus mRNA for high-affinity carnitine transporter,
ACCESSION	AB017260
VERSION	AB017260.1 GI:3869708
KEYWORDS	high-affinity carnitine transporter; CT1.
SOURCE	Rattus norvegicus (strain:Sprague-Dawley) adult intestine cDNA to mRNA.
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 3038) Sekine,T. Direct Submission Submitted (26-AUG-1998) Takashi Sekine, Kyorin University, School of Medicine, Department of Pharmacology and Toxicology; 6-20-2 Shinkawa, Mitaka, Tokyo 181, Japan (E-mail:OZU0742@nifty.ne.jp, Tel:0422-47-5511(ex.3451), Fax:0422-79-1321)
REFERENCE	2 (sites) Sekine,T., Kusunohara,H., Utsunomiyama-Tate,N., Tsuda,M., Sugiyama,Y., Kamali,Y., and Endou,H. Molecular cloning and characterization of high-affinity carnitine transporter from rat intestine Biochem. Biophys. Res. Commun. 251 (2), 586-591 (1998)
TITLE	location/Qualifiers
JOURNAL	1..3038
MEDLINE	/organism="Rattus norvegicus"
FEATURES	/strain="Sprague-Dawley" /db_xref="taxon:10116" /tissue_type="intestine" /dev_stage="adult" 101..1774 /gene="CT1" 101..1774 /codon_start=1 /product="high-affinity carnitine transporter" /protein_id="BAAI4399.1" /db_xref="GI:3869209" /translation="MRDYDEVTAFLGEGPPRIIFLLSASIIPIINGFNGMSIVELAG TPEIRCIAPHTVNLSSAMRNHSHIPLEFKDRQYQSCRRYLATIANFSALGLEPGRD VDLEQLQENCCLDGMEYNKGVPLSTIVTEMDLVCKDKWAKPLFTSLFEVGLMGSPFS GTSLDRQRNNVFILTMWGOTGSFDLQFVNEMFMFVLVFLVGMOGISNVVAAYFG

[illegible]



Qy	948	gttatacccttaagtcctcccccagatgctcatctcccaaggagacgattttagagggcaagaggt	1007
Db	925	gtttatcccttaagtcctcccccagatgctcatctcccaaggagacgattttagagggcaagaggt	984
Qy	1008	gatacatccgaaggctgtcgcaaacgaatggagattgtgtgctctcccatatcttgaacc	1067
Db	985	gatacatccgaaggctgtcgcaaatccaattccaatggagattgtgtgctctcccatatcttgaacc	1044
Qy	1068	gaatgaattacaaagactaaattccaagaagcagatgcccaacaattctgatactgct	1127
Db	1045	cagtagagttacaaagactaaattccaagaagcagatgcccaacaattctgatactgct	1104
Qy	1128	tcgagaccggagatatccggatggtgtccatatgttccatatgtctgtgtgagacatatc	1187
Db	1105	ccgaacacgaataattcagagatcatatccatatgttccatatgtctgtgtgagacatatc	1164
Qy	1188	agtggagctattttggcttgccttgatactccctcaactctgatatgggagacattgtgtaa	1247
Db	1165	agtggagctattttggcttgccttgatactccctcaactctgatatgggagacattgtgtaa	1224
Qy	1248	ctgtctctcttcagcgatggttgaagttcccaagacatagtgtgtgacctgtgtgtgtgca	1307
Db	1225	ctgtctctcttcagcgatggttgaagttcccaagacatagtgtgtgacctgtgtgtgtgca	1284
Qy	1308	atatgttccccggcgctatccatctgagccacgacctctctctctgggtgagctgtctctc	1367
Db	1285	gcacctgccccggcgctatccatctgagccacgacctctctctctgggtgagctgtctctc	1344
Qy	1368	ctctaatgcaactgtgtlaccacccagactgtatatttggctacagctccctgtgtatgttgg	1427
Db	1345	cttcatatcagctgtgtlaccacccagactgtatatttggctacagctccctgtgtatgttgg	1404
Qy	1428	caagttttgagtaacagcgtgtgctcttccatagtctcaagtgtacaaagccggagctgtacc	1487
Db	1405	caagttttgagtaacagcgtgtgctcttccatagtctcaagtgtacaaagccggagctgtacc	1464
Qy	1488	cacagttgttgagaaacatctgtgtgtgagtgtagctcccaacgaatcccgctgggagacat	1547
Db	1465	cacagttgttgagaaacatctgtgtgtgagtgtagctcccaacgaatcccgctgggagacat	1524
Qy	1548	ccgtgtctccctcatctcglttaacctgtgtgtgctgaagaccgcttccctcaatctcat	1607
Db	1525	cctgtgtctccctcatctcglttaacctgtgtgtgctgaagaccgcttccctcaatctcat	1584
Qy	1608	ggggaagttcgacatctcccgacagcacaactctaacctgtgttctcccaagagacttgcgtac	1667
Db	1585	ggggaagttcgacatctcccgacagcacaactctaacctgtgttctcccaagagacttgcgtac	1644
Qy	1668	cccaactcccaagacacatctgacccagatgtctaagatctcaagaaatgaaacacagaanaac	1727
Db	1645	ccctctctcccgacacacatctgacccagatgtctaagatgtcaaaagaaatgaaacacagaanaac	1704
Qy	1728	tcgaagttacacaagaagtgttaaaagtgttcaagaagaagcccaacatctcttaaaagcac	1787
Db	1705	ccaaagcccaaaacaaagagacgcaaaaaagatgtgagaaagcccaacagcgtcttaaaagacac	1764
Qy	1788	agaccttcaacatcgctccacgataagggagaaacatgaaagagaa	1831
Db	1765	agaccttcaacatcgctccacgataagggagaaacatgaaagagaa	1808
RESULT	7		
AF111425		1881 bp	linear
LOCUS			
DEFINITION	Mus musculus sodium-dependent carnitine transporter (Ct) mRNA,		ROD 05-JAN-1999
ACCESSION	AF111425		
VERSION	AF111425.1		
KEYWORDS	house mouse,		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

[illegible]

468 tggctgggaagttcaagtcagtcgtctacgttccacattgtgacgagtgaaactgct 527  
|||||  
444 tggctgggaagttcaagtcagtcgtctacgttccacattgtgacgagtgaaactgct 503  
|||||  
528 gtgtgaagcaactgtgaagggcccccacacaaatctcctgtcttctggtggtgct 587  
|||||  
504 gttgaagatgtacgtgaaagcccccacacacacacacacacacacacacacacac 563  
|||||  
588 gggctcctcattcattcagggcagctgtgacagaggtttggccggaagaagtgtgct 647  
|||||  
564 gggctcctcattcattcagggcagctgtgacagaggtttggccggaagaagtgtgct 623  
|||||  
648 gacattggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctcctc 707  
|||||  
624 gacattggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctcctc 683  
|||||  
708 gttgtcgtgctgt 767  
|||||  
684 gttgtcgtgctgt 743  
|||||  
768 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 827  
|||||  
744 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 803  
|||||  
828 gtgcattattatgcatctgtgcatcagtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 887  
|||||  
804 gtgcattattatgcatctgtgcatcagtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 863  
|||||  
888 ctggcgaagtcgtcgt 947  
|||||  
864 ctggcgaagtcgtcgt 923  
|||||  
948 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1007  
|||||  
924 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 983  
|||||  
1008 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1067  
|||||  
984 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1043  
|||||  
1068 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1127  
|||||  
1044 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1103  
|||||  
1128 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1187  
|||||  
1104 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1163  
|||||  
1188 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1247  
|||||  
1164 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1223  
|||||  
1248 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1307  
|||||  
1224 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1283  
|||||  
1308 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1367  
|||||  
1284 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1343  
|||||  
1368 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1427  
|||||  
1344 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1403  
|||||  
1428 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1487  
|||||  
1404 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1463  
|||||  
1488 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1547  
|||||  
1464 gttcctcgtggaagcaagtcagtcagtcgtcctcctcctcctcctcctcctcctc 1523  
|||||

QY 1548 cctgtcctcctactcgttttactccttggtgctcagacagccttccctcctacattc 1607  
|||||  
DB 1524 cctgtcctcctactcgttttactccttggtgctcagacagccttccctcctacattc 1583  
|||||  
QY 1608 ggggaagtcgtacacatcctgacagacacacacacacacacacacacacacacac 1667  
|||||  
DB 1584 ggggaagtcgtacacatcctgacagacacacacacacacacacacacacacacac 1643  
|||||  
QY 1668 cccactccacagacacatcgtacagatgctgaagtcgaaggaaggaaggaagga 1727  
|||||  
DB 1644 cccactccacagacacatcgtacagatgctgaagtcgaaggaaggaaggaagga 1703  
|||||  
QY 1728 tccaagtcacacaggaagtcgaaggaaggaaggaaggaaggaaggaaggaaggaag 1787  
|||||  
DB 1704 cccaagtcacacaggaagtcgaaggaaggaaggaaggaaggaaggaaggaaggaag 1763  
|||||  
QY 1788 agccttcaacatcgtcctcagtaaggaaggaaggaaggaaggaaggaagga 1831  
|||||  
DB 1764 agccttcaacatcgtcctcagtaaggaaggaaggaaggaaggaaggaaggaag 1807  
|||||

RESULT 8  
ABO15800  
LOCUS ABO15800 1888 bp mRNA linear ROD 19-JUN-1999  
DEFINITION Mus musculus mRNA for OCTN2, complete cds.  
ACCESSION ABO15800  
VERSION ABO15800.1 GI:4126482  
KEYWORDS octn2; OCTN2.  
SOURCE Mus musculus (strain:C57BL 6J) adult kidney cDNA to mRNA.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
1 (sites)  
Nezu,J.,Tama,I., Oku,A., Ohashi,R., Yabuchi,H., Hashimoto,N.,  
Nikado,H., Sai,Y., Koizumi,A., Shoji,T., Yakada,G., Matsushita,T.,  
Yashino,M., Kato,H., Ohura,T., Tsujimoto,G., Hayakawa,J.,  
Shimane,M. and Tsuji,A.  
Primary systemic carnitine deficiency is caused by mutations in a  
gene encoding sodium ion-dependent carnitine transporter  
Nat. Genet. 21 (1), 91-94 (1999)  
99113835  
MEDLINE  
2 (bases 1 to 1888)  
JOURNAL  
Nezu,J.  
AUTHORS  
Submitted (24-JUN-1998) Jun-ichi Nezu, Chugai Research Institute  
for Molecular Medicine, Inc., Gene Search Program: 153-2 Nagai,  
Nihari, Ibaraki 300-4101, Japan (E-mail:nezu@climmed.com,  
Tel:81-298-30-6211, Fax:81-298-30-6270)  
JOURNAL  
Tel:81-298-30-6211, Fax:81-298-30-6270)  
FEATURES  
source  
1..1888  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL 6J"  
/db\_xref="taxon:10090"  
/tissue\_type="kidney"  
/dev\_stage="adult"  
60..1733  
/gene="octn2"  
/gene="octn2"  
60..1733  
/gene="octn2"  
/product="OCTN2"  
/protein\_id="BAA36590.1"  
/db\_xref="GI:4126483"  
/translation="WRDDEYVAFLEGWSPFORLIFPLISAIIPNGNGMSIYELAG  
TPEHRCIVPHRYNLSAMRNHISPLETDGGOVOKCKRRRLATIANSEGLBGRD  
VDLEDDESDCHDEYDKVPLSTIVYEMDLCKDKDKAPLTTSLPFGVLMKSPIS  
GOLSRFGKNVLFUTMGQTFSLQVSVNFEXFLVYVGAGQISNVAAVLG  
TEILKSRIILFATLGVCIIFYAFGRMPLPLAFTIRDMRLLLALTVPGVJGALMWF  
IPESPRMLISOGRIKEAEVIRKAKINGIYAPSTIFPSELQDINSTKPOLHHIYDL  
IFRNIRVITISIIIMLTIISVYGSLDHPNHDYVNCFLAAVEYAVYLAWL  
LLOYLPRRYSTISALFISGSLVLMQIVPSLFYSTLSTALVWVGEGTISAVSMYVYT  
ALIVTYVRNMGVGVSTASRSLGSLSTPYLYLADYDFELTYIMGSLITITAILITLF  
FPESFVPLPDTIDQMLRVKGIKQWIOISOTRMOKDEESPTLKSTAF"





QY	1608	gggaagctgcacatctctgaaagccatcttacccttctgtttctccagagagctcggtac	1667
Db	1597	gggaagctgcacatctctgaaagccatcttacccttctgtttctccagagagctcggtac	1656
QY	1668	cccaactccagaaacccatctgaccagatgcctgaagtgtaagaagtgtaaacagagaaanc	1727
Db	1657	cccttctcccaatgacattgaccagatgcttaagggtcctcaagaggaatataacagtggtgcaaat	1716
QY	1728	tccaagtcacacaaagatglttaaaagatggtcacaagaaagcccaatcctctaaagcac	1787
Db	1717	ccaaagccagacaaagaatgcacaaaagaatgctgaaagaaagcccaacagctcttaagagcac	1776
QY	1788	agccttctacatcgcttccagtgtaaggaagaaactgaagagaa	1831
Db	1777	agccttctacacacccctgctccagaaagcccaaaactgattggaa	1820
RESULT	10		
LOCUS	AB018436	2297 bp	mRNA linear ROD 21-DEC-2000
DEFINITION	Mus musculus mRNA for OCTN3, complete cds.		
ACCESSION	AB018436		
VERSION	AB018436.1	GI:4996130	
KEYWORDS	OCTN3.		
SOURCE	Mus musculus adult testis cDNA to mRNA.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.		
AUTHORS	1 (sites)		
TITLE	Tamai, I., Ohashi, R., Nezu, J., Sai, Y., Kobayashi, D., Oku, A., Shimane, M. and Tsuji, A.		
JOURNAL	Molecular and Functional Characterization of Organic Cation/Carboline Transporter Family in Mice		
PUBMED	J. Biol. Chem. 275 (51), 40064-40072 (2000)		
REFERENCE	11010964		
AUTHORS	2 (bases 1 to 2297)		
TITLE	Nezu, J.		
JOURNAL	Direct Submission		
PUBMED	Submitted (07-OCT-1998) Jun-ichi Nezu, Chugai Research Institute for Molecular Medicine, Inc., Gene Search Program, 153-2 Nagai, Nihari, Ibaraki 300-4101, Japan (E-mail: nezu@climmed.com, Tel:81-298-30-6211, Fax:81-298-30-6270)		
FEATURES	Location/Qualifiers		
source	1..2297		
gene	/organism="Mus musculus"		
CDS	/db_xref="taxon:10090"		
	/tissue_type="testis"		
	/dev_stage="adult"		
	171..1865		
	/gene="OCTN3"		
	171..1865		
	/gene="OCTN3"		
	/note="Carnitine transporter in testis"		
	/codon_start=1		
	/product="OCTN3"		
	/protein_id="BA078343.1"		
	/db_xref="GI:4996131"		
	/translation="MIDYDEVTAFLGFMGTGFORLIFPLLSAIIIPNGFTGISAVFLTAIPERCRIPDVNLSAMRNHSTPMERKDGPEYQKRRRLATIANPSELGEPGDVLDLSEQENLDGWEYKDFLFTSYIVTEMDLVCKDWMKAPLTTSPFYGVLLSFISGQSLDRGKCNKLEFLTMAMHGFSEFIOVESVNEFEMFLVTLVGMGHSNYVAVFVGLTMSLSCSVRIIFATIGVCIFPAFGFMVLPFAVYIRMRRLILALTPGYCGALMWFIPESPMILISGGRKEAEVIRKAKINGIYAPSTIDPESPTNKLODSSKPKSHHIIAMLLQPNIRLTITMSIIIMWTISVGEYGLDHPNNGNTYVNCPLIAVEVPAVYVYLLQHVSRYSRYSMAGSLFLGGSVLLVQLVPSDLHLSTLTVWYVGRGITSASMYVYVTELEPIVVRNMGVGSVSTASRLGSLISFYEYLDAYDRRLPYIIMGSLTITLITITITLTPPESSGVSLEPTIDEMQVKKLKQROSLSKSPKSKGNSVRSRTSEPKGF"		
BASE COUNT	523 a	571 g	584 t
ORIGIN			1 others
Query Match	63.9%;	Score 1170;	DB 10; Length 2297;
Best Local Similarity	82.0%;	Pred. No. 5e-203;	

	Matches	1363;	Conservative	0;	Mismatches	290;	Indels	9;	Gaps	1
QY	113	ctgagggcgcgcatactgacacgaggtgacacgcttctctgagcgagtggggccct	172							
Db	160	CTGGCGACTCATGCTTGAAGTACTAGACGAGGTGACCGGCTCTCTGGCGAGTGGGGACCT	219							
QY	173	tccagcgctcaatcttcttctctgcttaagcgcagcatcatcccaatggtctaacggcc	232							
Db	220	TCCAGCGCTCATCTTCTTCTCTCTCAGCGCCAGCATCATCCCAATGGCTTTACTGGGT	279							
QY	233	tgctctcggttctctgatatagcgaccccggaacacacgcttgcgcggtgtcgcgagccgca	292							
Db	280	TGTCAGCGGTGTTCTCTACGCGGCTATCCCGGACACCGTGGCGGATACCAACACCTGA	339							
QY	293	accgagcagcgccctgycgcgaacacacgcttcccaatcgcgctgcgcgagcgcgcgaag	352							
Db	340	ACCTGACGACGCGCTGGCGGCAACCAAGTATCCGATGTGAGACGAAAGGACGCGACCAAGG	399							
QY	333	ttgcccaacagctgcgcgcgtctacgcgctcgcacaatcgcacaattctcggcgctcggc	412							
Db	400	TGCCCTCAGAAATGCCCGCGCTTACCGATCGCATGCGCACCATCGCCAACTTCTGAGCTGGGGC	459							
QY	413	tggagccggggcgcgagctgagaccttgggcaactctggagcagaagagcgtctgtatgct	472							
Db	460	TGGAGCCGGGACGGGAGCTGGACCTGGACACTGGAGAGGAGAACTGCTGGATGGCT	519							
QY	473	ggaggtctcaagcagagcttacctctgtccacatttgtgacgaagtggaaacctgggtgtg	532							
Db	520	GGGACTACGACAAAGACATCTTCTCTCCACCATGTGACAGAGTGGGACCTGGTGTGTA	579							
QY	533	aggacgacctgggaagcccccaatcacaaatctcttgccttctgtgggtgtgtgttggct	592							
Db	580	AGGATGACTGGAAAGCCCCACTCACCACTCCCTTCTTACGTGGGTGTGCTTAAAGCT	639							
QY	593	ccctcatttccagggcaactctgtcgaagaagtttggccggaagaatgtgtgttctgtgaca	652							
Db	640	CTTCTATTTCCGGGACACCTCTCGAGACAGGTTTGGTCGACAGAAATVCTTGTTTTGACCA	699							
QY	653	tgggcaatgcagaacaggtctcagcttctcgaaatcttcgcgaagattttagatgtttg	712							
Db	700	TGGCATTGCACACCGGATTTCAGCTTATACAAAGTCTTCTCTGTGAACTTCAGATGTTTA	759							
QY	713	tcgtgtctgttcttctgttagagcattggcgcgaatctccaactatgttgcgaacatttgc	772							
Db	760	CTCTGCTCTATACCCTTGTGGTGAATGGGACATATATCCAACTAGCTGGCAGATTGTC	819							
QY	773	tggggacagaaatctctgcgaagtcagcttcgctataataatctctacgtttagagatgtgca	832							
Db	820	TGGGACAGGAATCTTTCACAGTCAAGTATATATATTCGCCACCTTAGAGACTTTGCA	879							
QY	833	tatttatagcatttggctatagtgctgcgaactgtttgcttacttaaccgaacgcgcgc	892							
Db	880	TATTTTGGCTTTGGCTTATGCTGTGCTGTGCTCTGTGTTCAATCTTCAAGAGATGGA	939							
QY	893	ggatgtctgtgtggtgcgtctgacgaatgcgcgggggtgtgtgtgcgttgcgtacactctgtgtgtca	952							
Db	940	GGAGGCTGCTGCTGGCAATTACTTTACAGGCGTGTGTGTGGGCTCTGTHGTHGTTCA	999							
QY	953	tcacctgagtcctcccgatggtcattctctcagggaacgatttgaagaagcagaagtgtatca	1012							
Db	1000	TCCCGAATCCCCACGATGGCTCATCTCTCCAAAGCCGGAATTAAAGGACAGAGGTGATCA	1059							
QY	1013	tccgcgaagctgcgaacgaatggagatgttgttgccttcaactctcttgaaccgaagtgt	1072							
Db	1060	TCCGCAAGCTGCCAAAAATCAATGGGATTTGTTCACACTTCCACTTCTTCATCCAAAGC	1119							
QY	1073	ag-----ttaaagacctaaagtctcaagaagaagcagagctcccaacaactcttgcgattc	1123							
Db	1120	AGACCAATTAATTTACAAGAGCATATGTTTCAAGAAGCCCAAGTCGACACACATTTATGATC	1179							
QY	1124	tgtcttgcgaacctggaataatccggaatgtgtcaacaatcatgtctcaataatgtctgtgatgaca	1183							
Db	1180	TGTTTCCGAATCCCAAAATTCAGATCTCTCAACCTCATGTCTTTAATCTCTGGCTGTGACA	1239							

QY 1184 taccagtggtatatttggtcttgatatacttaacttgatgggaacatttg 1243  
Db 1240 TATCAGTGGGCTATTTTGAGCTATCTCTGACACTCTTAATTGATGGGAACATCTAAG 1299  
QY 1244 tgaacgtctccttccaaagatggtgaaagccagcatatgtgtgctgctgctgac 1303  
Db 1300 TGAATGCTTCTTACTAGGGGGGCTGTGAAGTCCCAAGCCATATGTGCGGGCTGGCTGTGT 1359  
QY 1304 tgaatatttgccggcgctatccatgagcaatgccccttctctggtgagtgatcc 1363  
Db 1360 TCGAGCATGTGTCGCCGCTTATTCATGTCGCTGCTCCCTCTCCGCGGTGGCACTGTCC 1419  
QY 1364 tctctctgagcagctgagcccccagacttgatatttgctcaagtcagtcgtgagtg 1423  
Db 1420 TTCTCTTAGTACACTGGTGCTTACAGACTACATTACTTGTCTACTACCTGGTGATGG 1479  
QY 1424 tgggcaagtttgagtcagcagctccttccatgctgctacgtgtacacagcagcgtct 1483  
Db 1480 TGGGAGAGTTTGGATCACCCTTGCTACTGCTATGCTATGTATACAGAGCTGAGCTGT 1539  
QY 1484 atcccacatggtgaaacatggtgtgtggaagtcagctccacagcatcccgctgagca 1543  
Db 1540 ACCCCACTGTGTGCAAAACATGGGGTGGGGGTGAGCTCCACAGCATCCCGCTTGGCA 1599  
QY 1544 gcaatctgtccctactcctgcttactcctgctgctgctcagcagcgtctcctgctcattc 1603  
Db 1600 GCATCTGTCTCCCTACTTGTGTTATCTAGTGCTATGACCGCGCTTACTTATATCC 1659  
QY 1604 tcaatggaagctgacacatcctgacagccatcctcacttgcttctccagagaagcttcg 1663  
Db 1660 TCATGGAGAGTCAACCATCTGACAGCTATCATCATCTTATTTCTCCAGAGAGTTCG 1719  
QY 1664 gtaacccactccagacacctatgaccagatgctaaagtgaaagaaagaaagaa 1723  
Db 1720 GGGTTTCTCTCCAGAAACATTTGACGAGATCAAAAGAACTAAAGAAAGGC 1779  
QY 1724 aaactccaagtcacacaagatgttaaagaagtcgacaagaa 1765  
Db 1780 AATCCTTAGCAAGAAAGGTCCTCCGAAGAGTCAAGGAA 1821

RESULT 11  
HSP1MP 2166 bp mRNA linear PRI 06-JAN-2001  
LOCUS H.sapiens mRNA for putative integral membrane transport protein.  
DEFINITION Y09881  
ACCESSION Y09881  
VERSION Y09881.1 GI:12053560  
KEYWORDS integral membrane protein; transport protein; UT2h gene.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2166)  
AUTHORS Spritzzenberger,F., Gruendemann,D. and Schoenig,E.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2166)  
AUTHORS Gruendemann,D.  
TITLE Direct Submission  
JOURNAL Submitted (05-DEC-1996) D. Gruendemann, University of Heidelberg,  
Department Of Pharmacology, Im Neuenheimer Feld 366, 69120  
Heidelberg, FRG

FEATURES  
source  
location/Qualifiers  
1..2166  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Caki-1"  
159..1814  
/gene="UT2h"  
159..1814  
/gene="UT2h"  
/note="putative"  
/codon\_start=1

/product="integral membrane transport protein"  
/protein\_id="CAA71007.1"  
/db\_xref="GI:12053561"  
/translation="MRDYDEVIFLEMGEPFORLIFLISASIIPIINGFMSVFLAG  
TPERRCRPDPAANLSSARNNSPRLRGRREVPHSRYLATIANFALGLEPDR  
VDGLOLEOESCLDGEESODVLTAVTEMLNCEENMMVPLPTSLFVGLLEGSVS  
GOLSDHGRKNVTFATMAVOTGSPLOITSISMETVLFVITGMQOISNYVAFLTG  
TEILGSVRIETSTLGVCTFANGVLLPLFAFTIDRMMLLALTPVLYCPPLMF  
IPSPFMLSORFREDIEDTQAAKMNINAVPAVFDSEELNPLKQKATLDLFR  
TRNIAIMTMSLLMLATSVGFALSLDAPNLDGAYLNCFSALIEIPAYITAMLL  
RTLPRIITAAVLFMGGVLLFIOLVPVDFYFSLDGLVMGFRGITSFASMSYFPAE  
LYPLIVNMAVGVSTASRVGSIIAFYFVGLAYNMLYIWMGSJLVJLITLFFP  
ESLGMPLPELLEOMOKVPMFRSGKTRDSETEBENKVLITAF"

BASE COUNT 509 a 556 c 535 g 566 t  
ORIGIN

Query Match 58.4%; Score 1069.6; DB 9; Length 2166;  
Best Local Similarity 78.3%; Pred. No. 9e-185;  
Matches 1298; Conservative 0; Mismatches 354; Indels 6; Gaps 1;

QY 102 tctgtggcctctgagggcgagcatgagcactacagcaggtgagcgccttctgagcga 161  
Db 137 TTTCGAGCGCGCAGTGGGAAGCATCGGAGCATACGAGAGTGATCCTTCTGGGCGCA 196  
QY 162 gtgggggcccctccagcgccctcatcttcttcctgctcaagcgcaagatcaatcccaatg 221  
Db 197 GTGGGGGCCCTTCAGGGCCCATCTTCTCTCTCTAGGCGCAGCAATCCCAATGG 256  
QY 222 cttaaccgacctgctcctcctgctcctgatagcagcccggaagcagcgcgcgggtgccc 281  
Db 257 CTTCAGATGATATGTCAGTGCAGTTCCTCGGGGGAGACCCCGAGAGACCGCTGTGAGAGGCC 316  
QY 282 ggaagcgcgagcactgagcagcgctctgagcaacacactgtccacttgcgtctgggga 341  
Db 317 GGAGCGCGCGCAACTGTGAGCAGCGCCTGCGCAACAAAGTCCCGCTGGCGGTGGCGGA 376  
QY 342 cggcgcggaaggtgccccacagctgcgcgcgctacagcgtcgacacatcgccaactctc 401  
Db 377 CGGCGCGAGAGTCCCCACAGCTCAGCCGCTACCGGCTGCCACCACTTCTCTC 436  
QY 402 ggcgcgcggagctgagcgcggggcgagcgtgagcactgagcagcgtggagagagaagctg 461  
Db 437 GCGCCTCGGGCTGTGAGCGGGCGCGAGCTGTGACCTGGGACACTGGAGCAGAGAGAGCTG 496  
QY 462 tctgagtgccgaggtgacgtacagcagctcctccttccacattgtgagcagagtgaa 521  
Db 497 CTTGATGCGCTGGGAGTTCAAGCAGGAGCTTACTCTCCACCGTCTGACCGAGTGGAA 556  
QY 522 cctggtgtgagagcagcgtgagagggcccaactcaatctcctgttctctgtggtgt 581  
Db 557 TCTGTGTGTGAGGACAACTGGAAGTGCCCTTCACACCTCCCTGTCTCTGTAGGCGT 616  
QY 582 gctgtgtggtccttcatcttcaaggcagcgtgtcaagacaggttggccggaagaatgtct 641  
Db 617 GCTCCGCTGGCTCCCTCTGTCGGGCGAGCTGTGAGAGGTTTGGCAGGAAGCACTGTC 676  
QY 642 gtctgagcagtgagcagtgagagagcgtcaacttcccttccaatcttccgaagatt 701  
Db 677 CTTCGCAACATGGCTTACAGACTGCTTACGCTTCTCAAGATTTCTCCATCAGCTG 736  
QY 702 tgaagatgttgcgtctgttctccttgaagcatgagcagatcccaactatgtgc 761  
Db 737 GGAGATGTTCACTGTGTTATTGTGCATGCGGGCATGGGCAATGTCGAATATGTGCT 796  
QY 762 agcaatgttcctcgaggaagaatcttgcgaagtcagtcgtataatctctcagct 821  
Db 797 AGCCTTATCTAGCAAGAAATTTCTGGCAAGTCAAGTCTGATTAATTCCTCATAT 856  
QY 822 agagatgtcatatttatcatattgctacatgagtcagtcgacacttctgactactat 881  
Db 857 AGGAGTGTGCAATTTTTCAGAGTGGCTATATAGTGTGCTGCCACTGTGTACTTCAAT 916





Db 485 CTTGATGGCTGGAGATTGACGACGAGTCTACTGTCACACCTGCGTACGAGTGGAA 544  
Qy 522 cctggtgtgtgaagagacgtgaagagcccaactcaactcctcgtgtcctcgtggtgt 581  
Db 545 TCTGGTGTGTGAGGACCACTGGAGAGTGTCCCTCCACCTCCCTCTCTCTGTAAGCGT 604  
Qy 582 gctgtgtgtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 641  
Db 605 GCTCTCTCGGCT 664  
Qy 642 gctgtgtgtgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 701  
Db 665 CTTCAC 724  
Qy 702 tgaagt 761  
Db 745 GGAGATGTGCTACTGTTATTTGTCATGCTGGGATGGGATGGGATGGGATGGGATGG 784  
Qy 762 agcattgtcctgt 821  
Db 785 AGCCTTCACTAGGACACAGAAATTTCTTGCAAGTCAAGTCAAGTCAAGTCAAGTCA 844  
Qy 822 aggaagtgtcatatttcatgtcatgtcatgtcatgtcatgtcatgtcatgtcatgt 881  
Db 845 AGGATGTGACACATTTTTCAGAGTGGCTATATGCTGCTGCTGCTGCTGCTGCTGCT 904  
Qy 882 cagagacgt 941  
Db 905 CAGAGACGTGGGAGT 964  
Qy 942 ctgt 1001  
Db 965 GTGGT 1024  
Qy 1002 aggaagtgtcatatttcatgtcatgtcatgtcatgtcatgtcatgtcatgtcat 1061  
Db 1025 TGAAGATATCATCCAAAAAGCTGCAAAAAATGACACACACACACACACACACAT 1084  
Qy 1062 tgaacccgt 1121  
Db 1085 TGATTTCTGTGAG-----GAGCTTAATCCCTTAAGACAGCAAGAACTTTTCA 1138  
Qy 1122 tctgt 1181  
Db 1139 CCTGTTCAGACCTGGAGATATGTCATATGACCATATGCTTGTGCTGATGAGATGT 1198  
Qy 1182 catatcaggt 1241  
Db 1199 GACCTCAAGTGGTACTTGTCTGTCTGTGATGCTCTTAATTTACATGAGATGCC 1258  
Qy 1242 tgtgaactgtcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1301  
Db 1259 CCTCAACTGTTTCT 1318  
Qy 1302 gctgtcaatatttgcctcctcctcctcctcctcctcctcctcctcctcctcctc 1361  
Db 1319 ATTGCGAAGCTGCCAGGGCTTATATCATAGCTGAGTACTGTCTGCGGAGAGAGTGT 1378  
Qy 1362 cctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1421  
Db 1379 GCTTCTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 1438  
Qy 1422 gctgtcaatatttgcctcctcctcctcctcctcctcctcctcctcctcctcctc 1481  
Db 1439 GCTGCGAAGATTTGGATCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1498  
Qy 1482 gtatcccaagtggtgagaaacatggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1541  
Db 1499 CTACCCCAACCTGTGTGAGGACATGCGGTGGGTCACATCCAGGCCCTCCAGAGTGG 1558  
Qy 1542 cagatcctgt 1601

Db 1559 CAGCATATTGCCCCCTACTTTGTTACTCTGCTGTCTACACAGATGCTGCTTACT 1618  
Qy 1602 tctcatgtggaagtcctgaccatcctcctgacacccatcctcctcctcctcctcctc 1661  
Db 1619 CGTATGGGTAGTGTGACTCTCTGATGGAATCTTACACCTTTTTCCTGAAAGTTT 1678  
Qy 1662 cgttaccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1721  
Db 1679 GGGATGACTCTTCTCAGAAACCTTAAAGCAGATGCAAGAAATGTTGATGATCTG 1738  
Qy 1722 aaaaactccaagtcacacaggaatgttaaaagtgtc 1759  
Db 1739 GAAAAAACAAGACTCATATGAGACACAGAAATC 1776

RESULT 13  
AF169831  
LOCUS  
DEFINITION  
Rattus norvegicus organic cation transporter OCTN1 (Octn1) mRNA.  
ACCESSION  
AF169831  
VERSION  
AF169831.1 GI:5679325  
KEYWORDS  
SOURCE  
Norway rat.  
ORGANISM  
Rattus norvegicus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
1 (bases 1 to 2258)  
Wua, X., Georgeb, R. L., Huang, W., Wanga, H., Conway, S. J.,  
Leibach, F. H., and Ganapathy, V.,  
Structural and functional characteristics and tissue distribution  
pattern of rat OCTN1, an organic cation transporter, cloned from  
placenta  
Biochim. Biophys. Acta 1466 (1-2), 315-327 (2000)  
20286310  
MEDLINE  
2 (bases 1 to 2258)  
Wu, X., Wang, H., Leibach, F. H., and Ganapathy, V.,  
Direct Submission  
Submitted (17-JUL-1999) Department of Biochemistry and Molecular  
Biology, Medical College of Georgia, 1120 15th Street, Augusta, GA  
30912-2100, USA

FEATURES  
source  
1..2258  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
1..2258  
gene  
/gene="Octn1"  
72..1733  
CDS  
/codon\_start=1  
/product="organic cation transporter OCTN1"  
/protein\_id="AAD46922.1"  
/db\_xref="GI:5679326"

/translation="MRDDEVIAFGLDMGPRORLIFLLASIIIPNGNGSVYFLAG  
TPEHCIVPHVNLSSAARMNSIPLEKDGQVOSRRRLATIANFSALGLEPGD  
VDIOLBOESCLDMEYSKDYFSTIYEMLVNVEDPKRPITLSLPEVYLCSFYS  
GOLDRFRKRVLETRIMAVQTFSEFVOIFSTNMFVLAIVMGQISNVVAFTIG  
TEILSKSVRLFSTLVCCTFFAIGVWVLEFAYIRMRMLALLTLPLGLCVPLWF  
IPESRMLISQREBEAEOIIOKAKKNNGIAPVVIDPELDELNSLKQOXYLIDL  
FKTRINATIVASVLMMLNLSVGFALSNVPLNHLGVLYNCFSLGLEVAPYFAML  
LRLTPRLRYIAGVLFMGGVLLVOVPEEDYNFVSGIWLNGKFGVTSFMSMYFT  
ALVPLTVRNNAVGITSMASRVSGIIVAPFYICAVARILPYILMSGLVILIGITLF  
FPESFGVLPENLEOMOKVRFRCGKKSIVSMDEENPKVILITAF"  
BASE COUNT  
558 a 565 c 566 g 569 t  
ORIGIN

Query Match 53.2%; Score 973.2; DB 10; Length 2258;  
Best Local Similarity 74.2%; Pred. No. 3.1e-167;  
Matches 1230; Conservative 0; Mismatches 428; Indels 0; Gaps 0;  
Qy 97 cgcgtctgt 156







member 4, clone MGC:18328 IMAGE:3670138, mRNA, complete cds.  
BC010590  
VERSION BC010590.1 GI:14714870  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2277)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (10-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
REMARK Contact: MGC help desk  
COMMENT Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [villalonbcm.tmc.edu](mailto:villalonbcm.tmc.edu)  
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Hollaway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAC Plate: 23 Row: 9 Column: 14  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4126604.  
FEATURES  
source  
location/Qualifiers  
1..2277  
/organism="Mus musculus"  
/db\_xref="LocusID:30805"  
/db\_xref="taxon:10090"  
/clone="MGC:18328 IMAGE:3670138"  
/tissue.type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."  
/clone\_11b="NC1\_CGAP\_Mam5"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
115..1776  
/codon\_start=1  
/product="solute carrier family (organic cation transporter), member 4"  
/protein\_id="AAH10590.1"  
/db\_xref="GI:14714871"  
/translation="MRDYDEVIAFLGEMQFORLLIFLILASIIPIINGFNGSVFLAG  
TPEHRLVDPDVNLISMRNHSIPLETGRDGYOVSRRYRLATIANSSAKLEGGD  
VLEIOLEOESCIDGMEYXKDILSTIVBMNLVCEDDMKTPILTSLEFVGLGSEFVS  
GGLSDRFGRKVFANMAVQISFVSQJESTIMWENFTVLFALVNGOISNVVAFILG  
TEILSKSVRIIFSTLGVCFEFAIGVMVLEFAYFIRDMRLALATLGLGKPLMWF  
IPESPMILISORFAEBOIIOKAKMSIVAPACIPIPLELOEINLSKOOKVILILD  
FRTNIAITTTVAVMIMLMTISYGFALSNLMDGDIYNLFGLILEVAPYFTML  
LRLTPRRYITAGVLFWGGGVLLIOVPEDEVNPYSIGVMKRGKGTISASMLYPT  
AELPYIVNMAVGTISNASRGTILAPFTYVLGAYNRLLPYILILGLLIGITLTF  
PFESFVTLPELLEDMQKVGFRGCKKSTVSDRESRPVLTAA"  
BASE COUNT 555 a 578 c 565 g 579 t  
ORIGIN  
Query Match 52.8%; Score 966.2; DB 10; Length 2277;  
Best Local Similarity 73.2%; Pred. No. 3.7e-166;  
Matches 1238; Conservative 0; Mismatches 453; Indels 0; Gaps 0;  
QY 69 aagccgcgcgcgttcccaagcccccagccgcgcgtctcgtgagcgcgcatgcs 128  
DB 60 AATCCTTTCTCTCCACCCACCTCCCTCACCGCAAGCTGAGAGAGAGAGTGGC 119

QY 129 ggactacagagatgaccgctctctctgagcgaagtgggggccccttccagccctcatctt 188  
DB 120 GGACTACGACGAGGTGATGCTCTTCTGGGCGAGTGGGGCCCTTCCAGCGCTCATCTT 179  
QY 189 ctctccgtctcagcgccagcatatccccaatgtcttccacgcgctctctcgttctct 248  
DB 180 CTTTCTTGTTCAGCGCCAGCATCATCCCAATGGCTTCAATGTTATGTCAGTGTCTCT 239  
QY 249 gataagacccccggagacacgcttgcgcgggtgcgcgaacccgaaactgagacgcctg 308  
DB 240 GCGGGGACCCCGGACACCGTGGTCTTCTACACTGTGAACCTGAGAGCTCCTG 299  
QY 309 ggcgaacacacatgtcccaatcgtcggtcgagcgcgcgagtggtcccaacgtgcgcg 368  
DB 300 GCGCAACGACACATCCCTTGGAGACGAGGACGACAGAGTGGCTTCAGAGCTGCGC 359  
QY 369 ccgctacccggtcgcacacatcgcgaacttctcggcgctcgggctgagaccgagcgca 428  
DB 360 CCGCTACCGACTGGCCACCATGCGCAACTTCTTGCGATGGGGCTGAGCCAGGACAGA 419  
QY 429 cgtggaacctggggcagcgtcgagcaggaagcgtctgtgagtctgaggatcagatcaga 488  
DB 420 CGTGAATCTGGAGCAGCTGGAGCAGAGAGCTGCGATGGCTGGGAGTACGACAAAGA 479  
QY 489 cgtctaacctgtccacacatgtgacccagtggaacctggtgtgtgagagcagctgaaagc 548  
DB 480 CATCTTCCCTGCTCACCATCTGACAGATGAGATGGTGTGTGAGGATGACTGGAAAC 539  
QY 549 cccactcaacatctctctgttctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 608  
DB 540 ACCCCTCACACCTCTCCCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 599  
QY 609 gctgtcagacagatgttggcgggaagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 668  
DB 600 GCTGTGACAGAGTTGGCGAGGAAGAGCTCTTTCACCAACAGGCTGTGACAGACGG 659  
QY 669 ctccagctccctgcagatctctcgaagaatttgaatgttctgtcgtgcgttctgtctct 728  
DB 660 ATTCAGCTTCGTCGAGATTTTCTTCAACCAACGAGGAGATGTTCACTGTGTGTGTGT 719  
QY 729 tgttagcatgagcagatctccaaactatgtgcagcaattgtcctgtggagcagaatctc 788  
DB 720 TGTGGGAGTGGGCCAGATCTCCAACTAGGTGTGGCTTTCATACGAGAACTGAATCT 779  
QY 789 tggcagatcagtgatataatctctcagcttgagagtgatgcatatttatgtcattgg 848  
DB 780 GAGCAAGTCGGTTCGATCATCTTCTCCACAACTAGAGACTGTACATTTTGGCAATCG 839  
QY 849 ctacatgtgtcgcacacgttcttacttacttccaggaactggcgagatgtcgtgtgagc 908  
DB 840 CTACATGTGCTCCGCGCGCTGTTTGCACTTTCATGAGAACTGAGAGATGCTGCTGTC 899  
QY 909 gctgaagatgccgggggtgtgtgtgcgtgcgcacatcgtgtgttcatccctgtgaccccc 968  
DB 900 CTTGACACTGCTGGCGCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 959  
QY 969 atggctcatctcctggaggaagatttgaagaggcagagtgatcatcgcgaaggtgcgaa 1028  
DB 960 GTGGCTGATTCCTCCAGAGAGAGATTTTGGAGAGGCCCAACAGATCATCCAGAAAGCCG 1019  
QY 1029 agccaatggatgt 1088  
DB 1020 GATGAACAGCATCGTGGGCGCAGCAGGAGATATTCGATCTCTAAGCTACAGAGGCTTAA 1079  
QY 1089 ttccaagaagcagcagatcccaacaatcttgatctgtcttgcgaacctggaatcagat 1148  
DB 1080 CTCTTGAAGCAGCGAAGATGATATCTGTGACCTGTTCAGAGCTGGAACATTTGCCAC 1139  
QY 1149 gttacacatcatgtccataatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1208  
DB 1140 CATTAACGCTGATGGGT 1199



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 17, 2002, 01:51:58 ; Search time 596.09 Seconds  
(without alignments)  
5273.818 Million cell updates/sec

Title: US-09-521-195b-4

Perfect score: 1831  
Sequence: 1 cggacgacgtcttggtgcgcct.....aggagagaactgaagagaa 1831

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq\_032802: \*  
1: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: \*  
2: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: \*  
3: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: \*  
4: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: \*  
5: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: \*  
6: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
7: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: \*  
8: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: \*  
9: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: \*  
10: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: \*  
11: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: \*  
12: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: \*  
13: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: \*  
14: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: \*  
15: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
16: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: \*  
17: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: \*  
18: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: \*  
19: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: \*  
20: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: \*  
21: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: \*  
22: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: \*  
23: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1831	100.0	1831	20	AAAX26880
2	1831	100.0	1831	21	AAAO9889
3	1831	100.0	3261	23	AAAS67216
4	1352.6	73.9	1888	20	AAAX26902
5	1352.6	73.9	1888	21	AAAO9890
6	1170	63.9	2297	21	AAAB8053
7	1066.4	58.2	2135	20	AAAX26879
8	966.2	52.8	2083	20	AAAX26898
9	516	28.2	700	22	AAH92241

10	516	28.2	25871	21	AAA09888
11	334.6	18.3	700	22	AAH92757
12	221.2	12.1	700	22	AAH92756
13	217.2	11.9	700	22	AAH92200
14	215.2	11.8	2533	23	ABLI0925
15	215.2	11.8	4533	23	ABLI0924
16	205.8	11.2	2070	21	AACT7247
17	205.8	11.2	2110	22	AAK51500
18	200.4	10.9	1901	21	AAZ44679
19	199.8	10.9	1950	22	AAAC85821
20	199.2	10.9	2501	22	AAAC85820
21	198.2	10.8	1662	19	AAV49559
22	198.2	10.8	1888	19	AAV49558
23	195.2	10.7	1697	22	AAAC83979
24	189.4	10.3	700	22	AAH92196
25	183.2	10.0	2493	21	AAAT75034
26	183.2	10.0	2502	21	AAAT75035
27	183.2	10.0	2966	22	ABAO9274
28	177.4	9.7	700	22	AAH92210
29	168	9.2	2455	22	AAAD17476
30	167	9.1	2054	22	AAH90049
31	166.2	9.1	1821	21	AAAB6204
32	166.2	9.1	2112	21	AAAB62014
33	163.4	8.9	1865	22	AAH90101
34	161.2	8.8	2087	23	ABLI23075
35	161.2	8.8	4087	23	ABLI23074
36	160	8.7	700	22	AAH92213
37	158.8	8.7	2123	21	AAZ29300
38	158.8	8.7	2127	22	AAAC85819
39	158.4	8.7	481	19	AAV60034
40	155.8	8.5	2171	20	AAV79585
41	155.2	8.5	370	19	AAV60033
42	154.2	8.5	1882	17	AAAT08702
43	154	8.4	700	22	AAH92199
44	154	8.4	700	22	AAH92242
45	151.6	8.3	2179	21	AAAI1146

ALIGNMENTS

RESULT 1	
AAAX26880	AAAX26880 standard; DNA; 1831 BP.
AC	AAAX26880:
XX	23-JUN-1999 (first entry)
DT	DNA encoding a protein with cation transporting activity.
XX	
DE	Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
XX	heart disease; cancer; anti-tumour drug; anticancer drug; ss.
KW	Homo sapiens.
XX	
OS	
XX	
FT	Key
FT	Location/Qualifiers
FT	124..1797
FT	/*tag= a
PN	WO9913072-A1.
XX	
PD	18-MAR-1999.
XX	
PF	07-SEP-1998; 98WO-JP04009.
XX	
PR	20-MAY-1998; 98JP-0156660.
XX	08-SEP-1997; 97JP-0260972.
XX	
PA	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX	
PI	Nezu J, Oku A:
XX	

DR	WP1: 1999-215062/18.
DR	p-PsDB; AAY01650.
XX	
PT	Genes homologous with organic cation transporters OCT1 and OCT2,
PT	useful in design of new drugs for treatment of diseases due to
PT	abnormality of the transporter functions
XX	
PS	ClaIm 2; Page 56-61; 97pp; Japanese.
XX	
CC	The present sequence encodes a protein with cation transporting
CC	activity. The genes are significantly homologous with organic cation
CC	transporters OCT1 and OCT2. The genes may used in drug development,
CC	particularly in the treatment of diseases due to abnormality of the
CC	organic cation transporter functions e.g. fatty liver, heart diseases
CC	and cancers, by controlling such as by inhibition or activation.
CC	Administration of anti-tumour and anticancer drugs in combination
CC	with a transporter protein inhibiting agent allows the agents to penetrate
CC	into the diseased cells to enhance the drug action.
XX	
XQ	Sequence 1831 BP; 348 A; 527 G; 513 G; 443 T; 0 other;

Query Match	100.0%;	Score 1831;	DB 20;	Length 1831;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1831; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

[illegible]

QY	721	ttgttcctgtgtgagcgtatggtggccaagatctccaactatgtgagcagatctgtctctgggaca	780
Db	721	ttttgttcctgtgtgagcgtatggtggccaagatctccaactatgtgagcagatctgtctctgggaca	780
QY	781	gaatatcttggcaagtcagtcgtataaatctctcaagctttagggatgtgcataatttat	840
Db	781	gaatatcttggcaagtcagtcgtataaatctctcaagctttagggatgtgcataatttat	840
QY	841	gcatttgcctacaatgtgtctgccactgttttgcttacttcaaccgagatctgcgagatctg	900
Db	841	gcatttgcctacaatgtgtctgccactgttttgcttacttcaaccgagatctgcgagatctg	900
QY	901	ctggtggtgcctagcagatgctgcgggggtgctgtgcgttgagcaatctgtgtgttcaatccctgag	960
Db	901	ctggtggtgcctagcagatgctgcgggggtgctgtgcgttgagcaatctgtgtgttcaatccctgag	960
QY	961	tcctccccaatggtcctcatctctcagagagcagatttgaaagagcgaggtgatactccgcagag	1020
Db	961	tcctccccaatggtcctcatctctcagagagcagatttgaaagagcgaggtgatactccgcagag	1020
QY	1021	gcttgcacaagccaatgggagtgtgtgtgtccttccactatcttgaccggagtgatgataca	1080
Db	1021	gcttgcacaagccaatgggagtgtgtgtgtccttccactatcttgaccggagtgatgataca	1080
QY	1081	gacctaaagttcccaagaaagacgacagctcccaacaatcttgatctgtctcgaaccctggagat	1140
Db	1081	gacctaaagttcccaagaaagacgacagctcccaacaatcttgatctgtctcgaaccctggagat	1140
QY	1141	atcccgatggttccacatcatgtctccataatgctgtgagatgacatatccagtgtgatacttt	1200
Db	1141	atcccgatggttccacatcatgtctccataatgctgtgagatgacatatccagtgtgatacttt	1200
QY	1201	gggcttctgcctgatactcctaactctgacatggggagacatcttgtgaactgcctcttca	1260
Db	1201	gggcttctgcctgatactcctaactctgacatggggagacatcttgtgaactgcctcttca	1260
QY	1261	gcgatggttgaagtcctccagacatagttgttgagctgcgtctgtcgtacaaatatttgccggg	1320
Db	1261	gcgatggttgaagtcctccagacatagttgttgagctgcgtctgtcgtacaaatatttgccggg	1320
QY	1321	gcgtattccatgtgcacatgcgcctctctctctgggtggcagtgctcctctcttcaatgagctg	1380
Db	1321	gcgtattccatgtgcacatgcgcctctctctctgggtggcagtgctcctctcttcaatgagctg	1380
QY	1381	gtacccccagactgttatatttggcttaagctccctcgatggtggtgaggttggtgagct	1440
Db	1381	gtacccccagactgttatatttggcttaagctccctcgatggtggtgaggttggtgagct	1440
QY	1441	agcgctgtccttttccatgtgtctcaagtgtacacagccgagctgtatcccaagttgtgaga	1500
Db	1441	agcgctgtccttttccatgtgtctcaagtgtacacagccgagctgtatcccaagttgtgaga	1500
QY	1501	aacatggtgtgtggggtcaagctccacacgaatcccgctcgggagagatctcttctccctac	1560
Db	1501	aacatggtgtgtggggtcaagctccacacgaatcccgctcgggagagatctcttctccctac	1560
QY	1561	ttcggttaacctgtgtgcctacgacgcgcttctctgcctcaatctcaatgtggaagtgcacc	1620
Db	1561	ttcggttaacctgtgtgcctacgacgcgcttctctgcctcaatctcaatgtggaagtgcacc	1620
QY	1621	atcccgagaagccatctctaacctgttttcccaaggagcttcggtaacctcccaatcccaagac	1680
Db	1621	atcccgagaagccatctctaacctgttttcccaaggagcttcggtaacctcccaatcccaagac	1680
QY	1681	accatttgcacagatctcgaagagtcacaaagaaatgaaacacagaaaacctccaagtacaca	1740
Db	1681	accatttgcacagatctcgaagagtcacaaagaaatgaaacacagaaaacctccaagtacaca	1740
QY	1741	aggatgttaaaaaagatgtgtcacaagaagccccaacatccttaaaagacacagctcttaacat	1800
Db	1741	aggatgttaaaaaagatgtgtcacaagaagccccaacatccttaaaagacacagctcttaacat	1800







```

|||||
Db 2503 CAGACAGGCTTACGCTTCCGAGATCTTCTCCAGAAATTTTGAATGTTTGCCTGCTG 2444
Qy 721 ttgtctcttgtaagcagatggcgagatctccaaataatgtagcattgttcctgggagaa 780
Db 2443 TTGTCTCTTGTAGGCAATGGGCGCAGATCTCCAACTATGTGGCAATTTGTCTGGGGACA 2384
Qy 781 gaaattcttgcaagtcagtcgtatataatctcttaagcttagagatgtgcatcttat 840
Db 2383 GAAATTTCTTGGCAAGTACGTTGGATATATTTCTTAAGCTTAGAGATGTGCATATTTTAN 2324
Qy 841 gcaattggtacatagtgtgctgcacatgtttgcttaacttaaccgagactgaggatgctg 900
Db 2323 GCATTTGGCTACATAGGTGCTGCACATGTTTGTACTTTCATCCGAGACTGGCGAGTCTG 2264
Qy 901 ctgtgtgctgagcagatggcgagggtgtgtgtgtgagcactggtgtgtatcccttgag 960
Db 2263 CTGTGTGGCTGAGAGATGGCGGGGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtGtG 2204
Qy 961 tccccccgagtcgcatctctcagggagcatttgaagagcagaggtgatcatccgcaag 1020
Db 2203 TCCGCCGAGTGGCTCATCTCTCAGGGAGCATTTGAAGAGGACAGAGTGCATCCGCAAG 2144
Qy 1021 gctgtcaaaagccaatgggagatgtgtgtgcttccactatcttgaacccagtgagttaca 1080
Db 2143 GCTGCCAAAGCCAAATGGAGTGTGTGCTTCCACTATCTTTGACCCGAGTGAGTTACAA 2084
Qy 1081 gacctaaagttccaaagagcagcagctcccaaatcttgatctgtcttgaaccttgaat 1140
Db 2083 GACCTAAGTTCCAAAGAAAGCAGAGCTCCACAAATTTCTGGATCTGCTTCAACCTGGAA 2024
Qy 1141 atccgagatggtcaccatcatgcatatgctgtgagatgagccatcaagtgagctattt 1200
Db 2023 ATCCGAGATGGTCAACCATCATGTCATATATGCTGTGGATGACATATCAGTGGGCTATTT 1964
Qy 1201 gggcttgcgttgatactccttaactgtgagtgaggaatcttgtgaactgtctcttca 1260
Db 1963 GGGCTTTGCGTTGATCTCTTAATCTTGCATGGGAGACATCTTTGTGAATCTTCTCTTCA 1904
Qy 1261 gsgatgttgaagttcccaagatatgtgtgtgagcgtgtgtgtgtgtgaatatgttcccg 1320
Db 1903 GCGATGTTGAAGTCCAGAGATATGTGTGCGCTGCTGCTGCAATATTTGCCCCGG 1844
Qy 1321 cgtatccatctgagcagatgcctctctctctgtgtgagtgatctctctctcatgagctg 1380
Db 1843 CGCATTTCCATGGGCAATGGCCCTCTCTGTGGGAGTGTCTCTCTCATGACAGCTG 1784
Qy 1381 gtacccccagaactgtatatttggctacagtcctcgtgtgagtgatggtggcaagttgagtc 1440
Db 1783 GTACCCCCAGACTGTATTTATTTGGCTACAGTCTCTGTGTATGTTGGGCAAGTTTGGAGTC 1724
Qy 1441 agggctgccttttccatgagctcagtgatagacagcgagctgfatcccaagtggtgaga 1500
Db 1723 AGGGCTGCTTTTCCATAGGCTCTACGTTACAGCCGAGCTGTATCCCAAGTGGTGA 1664
Qy 1501 aacatgggtgtgagagtcagctccacagacatcccgccctgggagcagcatcgtctccac 1560
Db 1663 AACATGGGTGTGGGAGTCACTCCACAGCATCCCGCTGGGAGCATCTGTCTCCCTAC 1604
Qy 1561 ttggttaacctgtgtcctacagcagcttccctgacctacattctcatlgtggaagtcagac 1620
Db 1603 TTCTTTTACCTTGTGCTCTACGACCGCTTCTGCTCCCTACATTTCTCATGGGAAAGCTGACC 1544
Qy 1621 atctctgacagcactcctcacctgtttctccagagagcttggttaccacacccagac 1680
Db 1543 ATCTCTGACAGCATCTCTCACTCTTCTCTCCAGAGAGCTTCGGTACCCCACTCCAGAC 1484
Qy 1681 accatitgaccagatgtcctaagagtcctaagagatgaacaacagaaactcccaagtcataca 1740
Db 1483 ACCATTTGACCGAGATGTCTAAGAGTCAAAAGGATGATAACACAGAAATACTCCAAAGTCAACA 1424
Qy 1741 aggatgttaaaagatgtgcaagaaaggccaacatcctttaaagcacagccttcaaat 1800
|||||

```

```

Db 1423 AGGATGTTAAAGATGCTCAAGAAAGGCCACAAATCTTAAAGCACAGCCTTCAACAT 1364
Qy 1801 cgtcccaagtaaggagaaactgaagagaa 1831
Db 1363 CGCTTCCAGTAAAGGAGAAACTGAAGAGGAA 1333

```

```

RESULT 4
AA26902
ID AAX26902 standard; DNA; 1888 BP.
XX
XX
AC AAX26902;
XX
XX
DT 23-JUN-1999 (first entry)
XX
DE DNA encoding a protein with cation transporting activity.
XX
KW Organic cation transporter; OCT1; OCT2; drug development; fatty liver;
KW heart disease; cancer; anti-tumour drug; anticancer drug; ss.
XX
XX
OS Mus musculus.
XX
XX
FH Key Location/Qualifiers
FT CDS 60..1733
FT /*tag= a
XX
PN W0913072-A1.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP04009.
XX
PR 20-MAY-1998; 98JP-0156660.
PR 08-SEP-1997; 97JP-0260972.
XX
XX
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Nezu J, Oku A;
XX
XX
DR WPI: 1999-215062/18.
DR P-PSDB: MAY01652.
XX
XX

```

Genes homologous with organic cation transporters OCT1 and OCT2, useful in design of new drugs for treatment of diseases due to abnormality of the transporter functions

Claim 2; Page 80-85; 97pp; Japanese.

The present sequence encodes a protein with cation transporting activity. The genes are significantly homologous with organic cation transporters OCT1 and OCT2. The genes may used in drug development, particularly in the treatment of diseases due to abnormality of the organic cation transporter functions e.g. fatty liver, heart diseases and cancers, by controlling such as by inhibition or activation. Administration of anti-tumour and anticancer drugs in combination with a transporter protein inhibiting agent allows the agents to penetrate into the diseased cells to enhance the drug action.

Sequence 1888 BP; 392 A; 509 C; 486 G; 501 T; 0 other;

Query Match 73.9%; Score 1352.6; DB 20; Length 1888;  
Best Local Similarity 85.6%; Pred. No. 0;  
Matches 1505; Conservative 0; Mismatches 254; Indels 0; Gaps 0;

```

Qy 73 ccgcgcgttcccaagaccagccgcgctctgtgtgacctgagggcggaatgctggagac 132
Db 9 ccaaggtgtccctctctccatacagggcgctgtgtgagagctgagagcgatgcyggag 68
Qy 133 tagcagaagtgatccgcttctctgtgagtgagtgagggcccttccagcgcctatctctc 192
Db 69 tagcagaagtgatccgcttctctagcgagtgagtgagggcccttccagcgcctatctctc 128

```

[illegible]

Db	1209	gtccacgacctatgctctgcccgcgtgctgtgtgtgaagtaacttgcgccgcgtgatattc	1268
Qy	1333	gacacgccccttccctgtgtgacagtttccctctctcatgacagctgtgtccccagac	1392
Db	1269	tggcgctgcgcctttccctgtgtgacagtgcttctctcttccatgacagctgtgtcctca	1328
Qy	1393	ctgtatttatttgcgtacagctccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1452
Db	1329	tgttttattctgttccacttccctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1388
Qy	1453	tccatgtgtctatgtgttaacacgctcgagctgtatctccacacagtgtgtgtgtgtgt	1512
Db	1389	tccatgtgtctatgtgttaacacagcttgagctgtatcccatgtgtgtgtgtgtgtgt	1448
Qy	1513	ggagttacgttccacacagatcccgccctgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1572
Db	1449	gggggttcagctccacacagatcccgccctgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1508
Qy	1573	gggtgcctacgacgcgttccctgcctacatcttcatalgtgtgtgtgtgtgtgtgtgt	1632
Db	1509	gggtgcctatgacgttccctgcctatatttccatgtgtgtgtgtgtgtgtgtgtgtgt	1568
Qy	1633	atccttaccctgttttctccacagagcttgcgtgtaccctccacagacattgtaccag	1692
Db	1569	atccttaccctgttctccctcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1628
Qy	1693	atgcctaaaggtcacaagatgaataacacagaaaactctcaagtgcacatgtgttaaa	1752
Db	1629	atgcctaaaggtcacaagatgaataacagtgcacaaatccacagaccagatgtgttaa	1688
Qy	1753	gatgtgtcagaagaagagccacacatcccttaagaagcacagcttcaacatcggttca	1812
Db	1689	gatgtgtgaagaagaagccacacagcttccataagaagcaagcttcaacacgcttca	1748
Qy	1813	gggagaagaactgtgaagaagaa 1831	
Db	1749	gcaaaaactgtatgtgaaa 1767	
RESULT 5			
ID	AAA09890	standard; cDNA; 1888 BP.	
AC	AAA09890;		
XX			
DT	05-JUL-2000	(first entry)	
XX			
DE	Mouse OCTN2 cDNA sequence.		
XX			
KW	Organic cation transportation; mouse; carnitine transporter protein;		
KW	OCTN2; diagnosis; systemic carnitine deficiency; mutation; gene therapy;		
KW	juvenile visceral steatosis; ss.		
XX			
OS	Mus musculus.		
XX			
PN	WO200014210-A1.		
XX			
PD	16-MAR-2000.		
XX			
PF	07-SEP-1999; 99WO-JP04853.		
XX			
PR	07-SEP-1998; 98JP-0252683.		
XX			
PA	(CHUG-1) CHUGAI RES INST MOLECULAR MEDICINE INC.		
XX			
PI	Nezu J, Oku A;		
XX			
DR	WPI; 2000-256966/22.		
DR	P-PSDB; AAY83930.		
XX			
PT	Systemic carnitine deficiency gene OCTN2 encoding part of organic		
PT	cation transporter, useful as diagnostic tool -		

XX Example 1; Page 51-57; 106pp; Japanese.  
PS This sequence represents the cDNA encoding the mouse carnitine  
XX transporter protein OCTN2. The corresponding human sequence (AA09889)  
CC can be used as a target for diagnosis of systemic carnitine deficiency  
CC by detecting the presence of mutations in the sequence, especially seen  
CC in the disease juvenile visceral steatosis (jvs). The wild type OCTN2  
CC gene can be used in the gene therapy of the disease state.  
XX  
SQ Sequence 1888 BP; 392 A; 509 C; 486 G; 501 T; 0 other;

Query Match 73.9%; Score 1352.6; DB 21; Length 1888;  
Best Local Similarity 85.6%; Pred. No. 0;  
Matches 1505; Conservative 0; Mismatches 254; Indels 0; Gaps 0;

QY 73 cgcgcgcgttccagaccagccgcgcgtctctgtgagcctctagggcgagcgagac 132  
DB 9 ccacggtgtcccttattcccatagcgcgctgtggaagctgagagcgacgtcgagac 68  
QY 133 tacgacgaggttgacgccttcctcgtggaagtgggggacctccagcgccctcattcttc 192  
DB 69 tacgacgaggttgacgccttcctcagcgagtgggggccctccagcgccctcattcttc 128  
QY 193 ctgtctcagcgccagcatccatcccaatggtctcaaccgacctgtccctcgtctctgata 252  
DB 129 ctgtctcagcgccagcatccatcccaatggtctcaatggtatgtccatcgtctcccgcg 188  
QY 253 ggcagcccgagacgcgcgtcgcgggtgbcggagccgcgcagacctgagacgcctcgagc 312  
DB 189 gggagcccgagacgcgcgtcgcgggtgbcggagccgcgcagacctgagacgcctcgagc 248  
QY 313 aaccacactgtccacacgcgcgtcgcggagcgccgcagagtgcccccaactcccgccgcgc 372  
DB 249 aaccacacgtatcccggttgagagacgagagcgagacgagatgctccagaaatccgcgcgc 308  
QY 373 taaccgctccgcacacatccgcacactctccgcgtcgcggcttgagaccggcgagacgtg 432  
DB 309 taaccgactgcccacacatccgcacactctctcgaagctagagcgccggcgagacgtg 368  
QY 433 gacctggcgacgctggagagcgagagactgtctggaatgctgagagctcaagcagacgtc 492  
DB 369 gacctggcgacgctggagagcgagagactgtctggaatgctgagagctcaagcagacgtc 428  
QY 493 taactgtccacacactgtgacacgagtggaacctggtgtgtaggaagcagacgtcgaagcccca 552  
DB 429 ttctctgtccacacactgtgacacgagtggaacctggtgtgtaggaagcagacgtcgaagcccca 488  
QY 553 ctcaacactctctgtctctcgttgggtgtgctgtctgagctccctcaattcagagcagctg 612  
DB 489 ctcaacactctctgtctctcgttgggtgtgctgtctgagctccctcaattcagagcagctc 548  
QY 613 tcagacagaggtttggcgagagaaatgtcgtctcgtgacatctgggacatggaacagacgtc 672  
DB 549 tcagacagaggtttggcgagagaaatgtcgtctcgtgacatctgggacatggaacagacgtc 608  
QY 673 agctctctgacagatctctcgaagaatttgagatggttgctgagctgttgcctgtga 732  
DB 609 agctctctgacagatctctcgttgaacttcgagaatgtttcaagtgctttttgtccctgtg 668  
QY 733 ggcagtgagcagatctccaaactatgtagcagcatctgttcctgggagagaagaattctgagc 792  
DB 669 ggcagtgagcagatctccaaactatgtagcagcatctgttcctgggagagaagaattcttcc 728  
QY 793 aagtcagtctgataataatctctcagcttagagagtgatgatatattatgacttggtac 852  
DB 729 aagtcagtctgataataatctctcagcttagagagtgatgatatattatgacttggtac 788  
QY 853 atggtgtgcgaactgttctgttactcatccgagacgtgagagtgacgtggtggcgctg 912  
DB 789 atggtgtgcgaactgttctgttactcatccagagacgtgagagtgacgtggtggcgctc 848

QY 913 acgattccgggggtgtgtgtcgtgagcactgtgtgttcatccctcgaatcccccagtcg 972  
DB 849 actgtgtccaggggtgtgtatgtgtgggtctctctgtgtttcatccctcgaatcccccagtcg 908  
QY 973 ctcaatctcagggagacatttgaagagcagagtgatcatcccgaaagctgtccaaagc 1032  
DB 909 ctcaatctcagggagacatttgaagagcagagtgatcatcccgaaagctgtccaaagc 968  
QY 1033 aatgagattgtgtgtccttccactatctttgaccggagtgatgtatcaagactaagttcc 1092  
DB 969 aatgagattgtgtgtccttccactatctttgaccgaagtgtatcaagactaagttcc 1028  
QY 1093 aagaagcagcagatcccaacacattctggaatctgtcttggaaccttggaatccgagtcg 1152  
DB 1029 aagaagcagcagatcccaacacattctggaatctgtcttggaaccttggaatccgagtcg 1088  
QY 1153 accatcgttccataatgtctgtgagatgacacatcagtgaggtatcttggccttcgtc 1212  
DB 1089 accatcgttccataatgtctgtgagatgacacatcagtgaggtatcttggccttcgtc 1148  
QY 1213 gatctcctaactgtgagtgagagacatctttgtgaactgtcttccttccagcgatgtgaa 1272  
DB 1149 gatctcctaactgtgagtgagagacatctttgtgaactgtcttcctcagcgcgatgtgaa 1208  
QY 1273 gtccagacatagtgtgtgtccgtgcgtgcgtgcgaatatcttgcggcgcgatattccatg 1332  
DB 1209 gtccagacatagtgtgtgtccgtgcgtgcgtgtgtgtgacgactatgtcccgcgatattccatg 1268  
QY 1333 gccactgtccctctcctctgtgtgagagtgctcctctctcattcagtcagctgtgaccacagac 1392  
DB 1269 gccactgtccctctcctctgtgtgagagtgctcctctctcattcagtcagctgtgtccacagaa 1328  
QY 1393 ttgtatatttgatgacaaatctctgtgtgagtgagtgaggaagtttgagatcagctgtccctt 1452  
DB 1329 ttgtatatttgatgacaaatctctgtgtgagtgagtgaggaagtttgagatcagctgtccctt 1388  
QY 1453 tccatgtctacgtgttacacagccagagctgtlatacccaagtggtgagaaacaatggtgtg 1512  
DB 1389 tccatgtctacgtgtgtacacagccagagctgtlatacccaagtggtgagaaacaatggtgtg 1448  
QY 1513 ggaagtacgtccacagacatcccgctgtggagacatcgtctcctactcgtttaacct 1572  
DB 1449 ggaagtacgtccacagacatcccgctgtggagacatcgtctcctactcgtttaacct 1508  
QY 1573 ggtgtccataagacgcttccctgtccctcaactctcagtggaagctgtacatcctcagacgc 1632  
DB 1509 ggtgtccataagacgcttccctgtccctcaactctcagtggaagctgtacatcctcagacgc 1568  
QY 1633 atcctcactgttctccccaagagactcgtgtacccaactccccaagacatctaacag 1692  
DB 1569 atcctcactgttctccccaagagactcgtgtacccaactccccaagacatctaacag 1628  
QY 1693 atgtctaagagtcacaagaaatgaacacagaaactccaagtccacaagaagtgttaaaa 1752  
DB 1629 atgtctaagagtcacaagaaatgaacacagaaactccaagtccacaagaagtgttaaaa 1688  
QY 1753 gatgtgtcaagaagggcccaactcctttaaagaacagcctttcaactcgtctccaagtaa 1812  
DB 1689 gatgtgtcaagaagggcccaactcctttaaagaacagcctttcaactcgtctccaagtaa 1748  
QY 1813 gggagaaactgaagaaggaa 1831  
DB 1749 gcaaaaaactgtattggaaa 1767

RESULT 6  
AAA88053  
ID AAA88053 standard; cDNA; 2297 BP.  
XX  
AC AAA88053;  
XX  
DT 11-DEC-2000 (first entry)  
XX

DE Mouse OCTN3 encoding cDNA SEQ ID NO:2.  
XX Mouse; transporter; OCTN1; OCTN2; OCTN3; organic cation transporter;  
KM Identification; regulator; carnitine transport; ss.  
OS Mus musculus.  
XX Key Location/Qualifiers  
FH 171..1865  
FT CDS /tag= a  
FT /product= "OCTN3"  
FT /note= "organic cation transporter"  
XX MO200046368-A1.  
PD 10-AUG-2000.  
XX 04-FEB-2000; 2000MO-JP00619.  
XX 05-FEB-1999; 99JP-0028406.  
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX Nezu J, Ose A;  
PI MPI: 2000-586982/55.  
DR P-PSDB; AAB20578.  
XX Organic cation transporter gene OCTN3 expressed in testis for  
PT Identification of regulators of carnitine transport for use as drugs  
XX  
XX Claim 1: Page 39-45; 58pp; Japanese.  
XX  
XX The present invention describes a mouse organic cation transporter  
CC protein (OCTN3). Also described are: (1) a method for screening  
CC compounds for their ability to regulate the transport of an organic  
CC cation into the cell, by generating a cell expressing OCTN3 at the  
CC cell membrane, contacting with the compound and organic cation, and  
CC observing the degree of transport of the organic cation; and (2) a  
CC method for screening compounds for their ability to be transported into  
CC the cell by OCTN3, by generating a cell expressing OCTN3 at the cell  
CC membrane, contacting with the compound and observing the degree of  
CC transport of the compound. OCTN3 can be used for the identification of  
CC regulators of the transport of organic cations (especially carnitine)  
CC into cells by OCTN3, for use as drugs. The present sequence encodes  
CC mouse OCTN3.  
XX  
XX Sequence 2297 BP: 523 A; 618 C; 571 G; 584 T; 1 other;  
SQ

Query Match 63.9%; Score 1170; DB 21; Length 2297;  
Best Local Similarity 82.0%; Pred. No. 2.2e-267;  
Matches 1363; Conservative 0; Mismatches 290; Indels 9; Gaps 1;

QY 113 ctgagggcgcatgaggaactacgaagcggagcgccttcctgagcagtgaggccct 172  
DB 160 ctgcgcgaactctatgtactacgaagcggagcgccttcctgagcagtgaggccct 219  
QY 173 tccagcgcctatctctctcgtcgcagcgcagcatatccccaatgtctcaaccgcc 232  
DB 220 tccagcgcctatctctctcgtcgcagcgcagcatatccccaatgtctctactggt 279  
QY 223 tgtctcgtgttctctgtatgacaccccgagcagcgcgtgcgggtgacgagccgga 292  
DB 280 tgtcagccgtgttctctgacgagcatcccgagcagcgtgtgcggtatcagacaacg 339  
QY 293 acctagacagcgtcgtgacgaacacactgttccactgcggtgcggagcgcgagag 352  
DB 340 acctagacagcgtcgtgacgaacacactgttccactgcggtgcggagcgcgagag 399  
QY 353 tgcacacagcgtcgcagcgtcgcagcagcagcagcagcagcagcagcagcagcagc 412  
DB 400 tgcctcagaataatgcgcgcgtacacgactgycacacatgcacactctctgagctgggc 459

QY 413 tggagccggggcgcgagcgtgagacgtggcagcttgagagaggaagcgttctgagatgct 472  
DB 460 tggagccggggcgcgagcgtgagacgtggcagcttgagagaggaagcgttctgagatgct 519  
QY 473 gggagctcaatcagagcgtctacactgtccacacactgtgacagagtgaaacctgtgtgt 532  
DB 520 gggagctcaatcagagcgtctacactgtccacacactgtgacagagtgaaacctgtgtgt 579  
QY 533 aggaacgttgaaagccccaactacaactctctgttctctgtgtgtgtgtgtgtgtgt 592  
DB 580 aggaatgacttgaaagccccaactacaactctctctcttcttcttcttcttcttcttct 639  
QY 593 ccttcattcaggcagcgtctcagacaggtttgtgcggaggaatgtgtgtgtgtgtgtgtgt 652  
DB 640 ccttcattcaggcagcgtctcagacaggtttgtgtgcgaagaatactgttttgacaca 699  
QY 653 tgggcatgacagagcgttcagcttcctcagagatctctctcgaagaatttggagatgtg 712  
DB 700 tggcattgacacacggatcagcttcatacaagcttctctgtgaacttcgagatgttta 759  
QY 713 tctgtcgtgttctccttctgtagcagatggccagatctccacatactgtgcagcatgttcc 772  
DB 760 ctctgctctataccctctgtgtgagatggagacatatacctaactcgtgcagcatgttcc 819  
QY 773 tgggacagagaattcttgcgaatcagctcgtatataattctcctaagcttaagatgtgca 832  
DB 820 tgggacagagaattcttgcgaatcagctcgtatataattctcctaagcttaagatgtgca 879  
QY 833 tatttatgacttggtctacatgtgtctgcagctgttcttactatccacagactgtgc 892  
DB 880 tatttttggttggcttctcatgtgtgtcctctgttcttcatcctatcacaagatgtga 939  
QY 893 ggaatcgtcgtgtggtgcgtgacatgacgaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 952  
DB 940 ggaatcgtcgtgtggtgcgtgacatgacgaggtgtgtgtgtgtgtgtgtgtgtgtgtgt 999  
QY 953 tccctgtatcccccagatggtctatctctcgaagagagatttgaagagagagatgtatca 1012  
DB 1000 tccctgtatcccccagatggtctatctctcgaagagagatttgaagagagagatgtatca 1059  
QY 1013 tccgcaagcgttcgcaagacgaatgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1072  
DB 1060 tccgcaagcgttcgcaagacgaatgagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1119  
QY 1073 ag-----ttacaagactaagttccaagaagcagcagctccacaacatctgtgac 1123  
DB 1120 agacacaataattacaagaagatagttccaagaagcagcagctccacaacatctgtgac 1179  
QY 1124 tgcctgaaccttggaataatcgcgaatgtgtcaacatcgttccatcattgtgtgatacca 1183  
DB 1180 tgcctgaaccttggaataatcgcgaatgtgtcaacatcgttccatcattgtgtgatacca 1239  
QY 1184 tatcagtgagcgtatcttggcgttgcgttgatatacttctcaactgtatggcggagcattttg 1243  
DB 1240 tatcagtgagcgtatcttggcgttgcgttgatatacttctcaactgtatggcggagcattttg 1299  
QY 1244 tgaactgtctcttccagcagatgttgaagttcccaacataatgtgtgtgcgtgtgtgc 1303  
DB 1300 tgaactgtctcttccagcagcgtgtgtgaagttcccaacataatgtgtgtgcgtgtgtgc 1359  
QY 1304 tgcataattgcccggcggtatcgaatgtgcacatgtccctctctcgtgtgtgagatgtcc 1363  
DB 1360 tgcagcatgtgtcccgcggtatctatagctgtgtccctctcctcgtgtgtgagatgtcc 1419  
QY 1364 tctcttcatgagcgtgtgtatccccagagcttgatatttggtatgataatcgtcgtgtgtatg 1423  
DB 1420 tctcttcatgagcgtgtgtatccccagagcttgatatttggtatgataatcgtcgtgtgtatg 1479  
QY 1424 tgggcaagtttggatgacagcgtctcttccatgtgtctactgttaccacagcagcagcgtgt 1483  
DB 1480 tggggaagtttggatgacacactctgtccactcactatgtgtctatgtgtacacagcagcagcgtgt 1539

Qy	1484	atcccaagtggtgagaaatggtgtgtgtgagtgtaagtcacagatcccgctgtggca	153
Db	1540	accaccactgtgtcagaataatggtgtgtgtgtcgaagctccacagatcccgcttggca	1599
Qy	1544	gcacccgtctccctactctgtttactctgtgtgacctaacgcgtctctgcctatactc	1603
Db	1600	gcacccgtctccctactctgtttactctgtgtgacctaacgcgtctctgcctatactc	1659
Qy	1604	tcatgtgaaagtctgacacatccctgacagcaccatccactgtttctcccaagagctcg	1663
Db	1660	tcatgtgaaagtctgacacatccctgacagcaccatccactgtttctcccaagagagttctg	1719
Qy	1664	gtaccccaaccccaagacacattgacacagatctaaagtgaaagaaatgaaacaagaa	1723
Db	1720	gggttctctccccaagaacattgacagagatgcaaaagtccaataaactaaacaagtc	1779
Qy	1724	aaactcaagtacacaagaagtgtttaagaatggttcaagaa	1765
Db	1780	aatccctaagcaagaagaaggttccccaagaaagtcttaagaa	1821

RESULT	7
AAx26879	
ID	AAx26879 standard; DNA; 2135 BP.

... AAX26879;

DT 23-JUN-1999 (first entry)

DE DNA encoding a protein with cation transporting activity.

KM Organic cation transporter; OCN1; OCN2; drug development; fatty liver;  
KM heart disease; cancer; anti-tumour drug; anticancer drug; ss.

OS Homo sapiens.

	Key	Location/Qualifiers
FH	CDS	147..1802
FT		/*tag= a
FT		

PN W09913072-A1.

PD 18-MAR-1999.

PF 07-SEP-1998; 98WO-JP04009.

PR 20-MAY-1998; 98JP-0156660.

XX

**XX:**

XX

DR P-PSDB; AAY01649.

PT Genes homologous with organic cation transporters OCT1 and OCT2

PT abnormality of the transporter functions

PS Claim 2; page 45-51; 97pp; Japanese.

CC The present sequence encodes a protein with cation transporting

CC transporters OCT1 and OCT2. The genes may used in drug development,

CC organic cation transporter functions e.g. fatty liver, heart diseases

## CC Administration of anti-tumour and anticancer drugs in combination with

CC into the diseased cells to enhance the drug action

Sequence 2135 BP; 499 A; 547 C; 530 G; 559 T; 0 other;

Query Match	58.2%;	Score 1066.4;	DB 20;	Length 2135;
Best Local Similarity	78.2%;	Pred. No. 7.7e-243;		
Matches 1296; Conservative	0;	Mismatches 356;	Indels 6;	Gaps 1.

[illegible]

QY	1122	ctcgtctcgaaacctggaatataccggagatcgagtcacacatcatgctccataatgctctgtgattgac	1161b
Db	1139	ctgtctcagagactcgatgaataatgaccataaagacattatgctcttgctcatgattgacgt	1168b
QY	1182	catatacaatggagctatttttgagcttgccttgcttgatatactctaactgcatggaggacattt	1241b
Db	1189	gaacctcagttggttaacttctgtcctctgcggtgtccccaatttacaagagaaagacctta	1258b
QY	1242	tgtgaactgcttcctttcaagcagabggtttgaagtcacacagatatgtgttgccctgctgct	1301b
Db	1259	ctgaacagcttctcctctcctctgccttgattgaattcagattcaactaatacagcctgctgct	1318b
QY	1302	gctgcgaatattggcccgagcgatataccagagcaactgccttccttccttcctgggtgagaggt	1361b
Db	1319	attcggaacagctgcgccagcggtatatacatagactgcagtaacgtctctcggggagagaggt	1378b
QY	1362	cctctctcttcagctagctgblacccccagactgtatattttgctctacagctctgtgatt	1421b
Db	1379	gctctctctctcatcaactggtacctgtgattatctattactctatcaatctgctgctcat	1438b
QY	1422	gttgaggcaagtttgtagtcaagcgtccttcttcacatgagtcctgaagtgatacaagcgcgact	1481b
Db	1439	gtctgggaaaatttgggtacacccctcgtcttcttcacatgctgattgtctctcaacgtctgacgt	1498b

XX	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA	
XX	
PI	Nezu J, Oku A;
XX	
DR	WPI; 1999-215062/18.
XX	P-PSDB; AA101651.
PT	
XX	
XX	Genes homologous with organic cation transporters OCT1 and OCT2,
XX	useful in design of new drugs for treatment of diseases due to
XX	abnormality of the transporter functions
PS	
XX	Claim 2; Page 68-74; 97pp; Japanese.
XX	
CC	
CC	The present sequence encodes a protein with cation transporting
CC	activity. The genes are significantly homologous with organic cation
CC	transporters OCT1 and OCT2. The genes may used in drug development,
CC	particularly in the treatment of diseases due to abnormality of the
CC	organic cation transporter functions e.g. fatty liver, heart diseases
CC	and cancers, by controlling such as by inhibition or activation.
CC	Administration of anti-tumour and anticancer drugs in combination with
CC	a transporter protein inhibiting agent allows the agents to penetrate
CC	into the diseased cells to enhance the drug action.
XX	
SQ	Sequence 2083 BP; 477 A; 549 C; 548 G; 509 T; 0 other;

Sequence 2083 BP; 477 A; 549 C; 548 G; 509 T; 0 other;

Query Match	52.88;	Score 966.2;	DB 20;	Length 2083;
Best Local Similarity	73.28;	Pred. No. 4.2e-219;		
Matches 128;	Conservative	0;	Mismatches 453;	Indels 0;
			Gaps	0;

RESULT	8	
AAAX26898		
ID	AAAX26898	standard; DNA; 2083 BP.
XX		
AC	AAAX26898;	
XX		
DT	23-JUN-1999	(first entry)
XX		
DE	DNA encoding a protein with cation transporting activity.	
XX		
KW	Organic cation transporter; OCT1; OCT2; drug development; fatty liver;	
KW	heart disease; cancer; anti-tumour drug; anticancer drug; ss.	
XX		
OS	Mus musculus.	
XX		
Key	Location/Qualifiers	
FT	CDS	122..1783
FT		/*lag- a
XX		
PN	MO9913072-A1.	
XX		
PD	18-MAR-1999.	
XX		
PF	07-SEP-1998;	98WC-JP04009.
XX		
PR	20-MAY-1998;	98JP-0156660.
PR	08-SEP-1997;	97JP-0260972.
PR		

[illegible]







[illegible][illegible][illegible]



Db 1713 tcgattggttctgtacacagtggtgcttaactacgagctgtgctgtggaacacacaacatctg 1772  
 QY 1229 atggggaacatcttctgtactctctcttcaagcagatggtgaagtcaccagatgtgt 1288  
 Db 1773 gtggaacacatctgttaacttaacttactctctgtgcgcggtgaatcccggtatatacgc 1832  
 QY 1289 tggcctggtctgtctgaataattgcccggcgctatccatgagccactgccccttcc 1348  
 Db 1833 tgcctcttctacttgaaccgtgggtcgtcgtctccatccctgtgcgtaacatgtatgg 1892  
 QY 1349 tgggttgacagtgctctctctctcaatgacgtgtaacccccaagctgtattattgcta 1408  
 Db 1893 tggccggaataatgactctgtcgtccacacatctctgtccgagacatgaattggtctg 1952  
 QY 1409 cagtcctgtgtatgttggaacagtttggaagtcacgagctgtcttccatgctgaactgt 1468  
 Db 1953 ttgctctgcgcataatgagaagctgtctattactctgtctatgagaaacatctactat 2012  
 QY 1469 acaacagcagctgtatcccaacagtggtgaacaaatggtgtggaagtcagctcaag 1528  
 Db 2013 tctcagaggaacagttcccgactgtgtgcgaatgtggtgtcgtgagagctctccatg 2072  
 QY 1529 catcccgctgggagacatctctctctactctgttaactctgtgtgcttaagacgct 1588  
 Db 2073 tggctcgtgtggtggtgcatctctgcaaccctcaacaaactgctggcgagatctggcgac 2132  
 QY 1589 tctgcccataatctcaatggaagctctgacatccctgacagccatccactgtttc 1648  
 Db 2133 cgtgcgcagtcgatcatctctgagagacactgtctcaccgctgtgcctgtctctc 2192  
 QY 1649 tcccaagagagcttcgtaccccaactcccaacacatctga 1688  
 Db 2193 tgcgagagacccttaacaacacatgctcgagacacatcga 2232  
 RESULT 15  
 ABL10924  
 ID ABL10924 standard; cDNA; 4533 BP.  
 AC ABL10924;  
 DX 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27254.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 OS Drosophila melanogaster.  
 PN WO2001/1042-A2.  
 PD 27-SEP-2001.  
 PE 23-MAR-2001; 2001WO-US09231.  
 PR 23-MAR-2000; 2000US-191637P.  
 PI 11-JUL-2000; 2000US-0614150.  
 PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PMD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR P-SDB; ABB66821.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Claim 1: SEQ ID NO 27254; 21pp + Sequence Listing; English.  
 SS The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB157737-AB172072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 4533 BP; 1170 A; 1056 C; 1045 G; 1262 T; 0 other;  
 Query Match 11.8%; Score 215.2; DB 23; Length 4533;  
 Best Local Similarity 49.8%; Pred. No. 3; 3e-41;  
 Matches 618; Conservative 0; Mismatches 598; Indels 24; Gaps 2;  
 QY 470 gcttgagatcagtcagcagctctacctgtccacatgtgacgagtggaacctgtgt 529  
 Db 1996 gctacgtttaagatcgcgagcaagatctcaatagtcggtgacgcgagtggaacctgtgt 2055  
 QY 530 gtagagacgactggaagccccaactcaaatctcctgttcttctgtggtgtgtctgtg 589  
 Db 2056 gtagtcgaagctgtcgaatgtccaccacagtgatctgtattcaatgctggtggtgtctgtg 2115  
 QY 590 gctccttcaatcgaagcagctgtagcaaggtttgcccgaagaatgtgtctgtctgca 649  
 Db 2116 gaagcttaactcttgccagatcgcagatccgacaagcttggaagcaagccacactctgtcct 2175  
 QY 650 ccattggcactgacagacaggtctacatctctcagatctctcagatctctcgaagaatttgagatgt 709  
 Db 2176 cactgtgtgttcaagcactcaatctcgcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2235  
 QY 710 ttgt 769  
 Db 2236 aacagatctcccgatgattgt 2295  
 QY 770 tcttgaggagacaagaattctgtcaagtcagtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 829  
 Db 2296 tcatgccttgagatgt 2352  
 QY 830 gcaattttatgacattggtctacatggt 889  
 Db 2353 agatgtctctcgcgt 2412  
 QY 890 ggcgagatgt 949  
 Db 2413 ggcgt 2472  
 QY 950 tcatccctgagtcctcccgatgctcatctcctcaaggaagattgaagaagcagaagtgga 1009  
 Db 2473 tcatcccgagatctgcgcgt 2532  
 QY 1010 tcatccgaagctgtgcgaagccaatgt 1068  
 Db 2533 tcatccgaagctgtgcgaagccaatgt 2592  
 QY 1069 -----agtgttacaagaagactaagtctccaagaagaagcagatctcc 1108  
 Db 2593 tgggtgagcaggtgtgcgaaagaagaagcagaagatgtgcgcgtccctccaacacagcgg 2652  
 QY 1109 acaacatctgatatctgtctgaaccttggaatatccgatatgtgcacatctgtccataa 1168  
 Db 2653 ccactgt 2712  
 QY 1169 tgcgt 1228  
 Db 2713 tgcattgt 2772  
 QY 1229 atggggaacatcttctgtactctctcttcaagagatggtgaagtcaccagatgtgt 1288  
 Db 2773 gtggaacacatctgttaacttaacttactgt 2832



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2002, 23:48:24 ; Search time 135.85 seconds  
(without alignments)  
3310.673 Million cell updates/sec

Title: US-09-521-195B-4

Perfect score: 1831

Sequence: 1 cgaacgctctggtcgcct.....aggagaactgaagagaa 1831

Scoring table: IDENTITY NUC

Searched: Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA:\*  
1: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/lna/PCTUS.COMB.seq:\*  
6: /cgn2\_6/prodata/2/lna/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	183.8	10.0	1885	3	US-08-501-572-5
2	183.8	10.0	1885	3	US-09-040-444-5
3	158.4	8.7	481	1	US-08-592-126-109
4	155.2	8.5	370	1	US-08-592-126-108
5	155	8.5	1882	3	US-08-501-572-4
6	155	8.5	1882	3	US-09-040-444-4
7	150.4	8.2	1896	3	US-08-501-572-6
8	150.4	8.2	1896	3	US-09-040-444-6
9	124.8	6.8	2102	2	US-08-647-397-1
10	53	2.9	7218	1	US-08-232-463-14
11	52.2	2.9	1794	5	PCT-US95-14418-3
12	52.2	2.9	1794	5	PCT-US95-15327-3
13	52.2	2.9	3048	5	PCT-US95-14418-1
14	52.2	2.9	3048	5	PCT-US95-15327-1
15	49.4	2.7	4403765	4	US-09-103-840A-2
16	49	2.7	1686	2	US-08-648-657-14
17	49	2.7	2504	1	US-08-073-384C-3
18	49	2.7	2504	1	US-08-254-359A-3
19	49	2.7	2504	1	US-08-483-043-3
20	49	2.7	2504	1	US-08-481-238-3
21	49	2.7	2504	2	US-08-471-066B-3
22	49	2.7	2504	2	US-08-484-956-3
23	49	2.7	2504	2	US-08-757-653-3
24	49	2.7	2504	2	US-08-599-491-3
25	49	2.7	2504	2	US-08-756-386-3
26	49	2.7	2504	2	US-08-823-516-3
27	49	2.7	2504	3	US-08-682-853A-3

28	49	2.7	2504	3	US-08-759-038-3	Sequence 3, Appl
29	49	2.7	2504	3	US-08-758-314-3	Sequence 3, Appl
30	49	2.7	2504	4	US-09-350-309-3	Sequence 3, Appl
31	49	2.7	2505	1	US-07-977-434-9	Sequence 9, Appl
32	49	2.7	2505	1	US-08-458-819-9	Sequence 9, Appl
33	49	2.7	2505	5	PCT-US91-07035-9	Sequence 9, Appl
34	49	2.7	2640	1	US-08-384-490-30	Sequence 30, Appl
35	49	2.7	2640	1	US-08-459-383-30	Sequence 30, Appl
36	48.8	2.7	328	2	US-08-647-397-3	Sequence 3, Appl
37	47	2.6	833	2	US-08-403-852D-3	Sequence 3, Appl
38	47	2.6	833	3	US-08-510-646B-3	Sequence 3, Appl
39	47	2.6	833	4	US-09-231-818-3	Sequence 3, Appl
40	47	2.6	5392	2	US-08-403-852D-1	Sequence 1, Appl
41	47	2.6	5392	3	US-08-510-646B-1	Sequence 1, Appl
42	47	2.6	5392	4	US-09-231-818-1	Sequence 1, Appl
43	46.6	2.5	20235	1	US-07-642-734C-3	Sequence 1, Appl
44	46.6	2.5	20235	3	US-08-439-009A-3	Sequence 3, Appl
45	44.8	2.4	1611	2	US-08-551-211-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-08-501-572-5  
; Sequence 5, Application US/08501572  
; Patent No. 6063623  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; APPLICANT: Gorboulev, Valentin  
; TITLE OF INVENTION: Transport of protein Which effects The  
; TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals,  
; NUMBER OF INVENTIONS: 6  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Finegan, Henderson, Farabow, Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/501,572  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toomey, Kimberlin M  
; REGISTRATION NUMBER: 35,391  
; REFERENCE/DOCKET NUMBER: 02481.1453-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1885 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-501-572-5

Query Match 10.0%; Score 183.8; DB 3; Length 1885;  
Best Local Similarity 49.6%; Pred. No. 2.8e-36;  
Matches 622; Conservative 0; Mismatches 612; Indels 21; Gaps 5;  
Qy 434 acctggcagctgagcagagagagctgtcgtgagctgagctcagtcagtcagctct 493

D	b		A C A G A C C C A C C T T G C C G C T G G G T G C C T T G C C A G A A T G G C G G G T G T - - - - A T A C A C G C	460
O	y	494	a c c t g t c a c a c a t g t g a c c g a g t g g a a c c t g t g t g t a g a a c g a c t g y a a g g c c c c a c	553
D	b	461	C G G G C T T T C A T G C A T G C A C T G A G T T C A A C C T G G T G T G T C T A C T C C H C N G G A A G C T G G A C C	520
O	y	554	t c a a a t c e c t g t t c t l o c t b g t g g t g y g c t g t g t g g t c c t t a c t t c a g g c a g c t g t	613
D	b	521	T C T T Y T A G T O C T G T T T A A T G A T C G G G G C C T T C T T T G G G C T C T C T G T T G G C T A C T T T G	580
O	y	614	c a g a c a g i t t g g c c g g a a a t g t g t g t c t g t g a c c a t g g a c t g a a c a g a c g t c t a	673
D	b	581	C A G A C A G T T T G G C C C T A A C C T G T G C T C C T G G A A C T G C T G S T T C A A C C G G G T G C G	640
O	y	674	g e t t c o t g c a g a t c t t c t g a a a a t t t g a g a t g t t g c g t g t g t c t g t g t t g c t g t a g	733
D	b	641	G C G T C T A T A G G C C T T C C C C A C T A C A T G C A T G C A T G C T T C C G C T T C C G C C T G C A G G	700
O	y	734	g c a t g g g c a a g a t c c c a a c t a t g t g g a g a t t g t c c t g g a c c g a a a t t c t t g a c	793
D	b	701	G C C T G G T A G C A A G G G A A C C T G A T G G C T G E C T A C A C C T T A A T C A A G A A T T T G T T G G C	759
O	y	794	a g t a g t t c g f a t a a t a t t c t c a c g t t a g g a g t g r a c a t t t a t g a t t g t g t a c a	853
D	b	760	- - T G G G G T C C A A G A A A G S G T G S G C A T A T A C A A T G S C C T T A C G G G G G G C T G G	817
O	y	854	t g t g t c t g c a c a t g t t g t c t a c t t a c c a g a c t g r c g a t g c t g c t g t g t g c g t g a	913
D	b	818	T G G G C T T A C C G G G C C T G C C T A C G C C T G C T C A C I G G G G C G G G T G A G C T G A G G C A G C T	877
O	y	914	c g a t g c c g a g u g t a c t g t g t g t g a c a c t g r t g r t a c c c t a g t c c c c c a t g y c	973
D	b	878	C C C T G C C A C C C T T C C T T C C T G C T C T A C T A C T A G G T G T G C G G A G T C C C C T G S T G G C	937
O	y	974	t c a t c t c t a g g a c g a t t t g a a g a g g c a g a g t a c a t c c g c a a g y c t c c a a g c a	1033
D	b	938	T G T T A T C A C A A A A A G A A C T G A A G C A T A A A G T A A T G A C C A C A T T C G C T A A A A G A	997
O	y	1034	a t g g a t t g t g c t t c c a c a t c t t t g a c c c g a g t a g t t a c a a a c c t a a g t t c a	1093
D	b	998	A T G G G A A G T T C C C T C T G C T G A T T T A - - - - A G A T C C T T T C C C T A A A G A G A T G T C A	1051
O	y	1094	a g a a g c a c a g t c c c a c a a c a t t c t g a t c t g t c t c t g a a c c t g s a t a c c g a t g t c a	1153
D	b	1052	C G G A A A A C T A G G C C T T C A T T T T G A G A C C T T T C C G C A C G C G C C T G A G A A G C G C A	1111
O	y	1154	c o a t a g t l o c a t a t g t g t g a t a g a c a t a c a t a c a t a g t g c a t t t g g t c t g c t g	1213
D	b	1112	C C T T A T C T C T A T F A C C T G T G G T T C A G S A C T T G T G C T T R A T C A G S G C T C A C T C T G C	1171
O	y	1214	a t a c t c t a a c t t g c a t g g g a a c a c t t t g a a c g t c t c t t c a g c a g a t g t t g a g	1273
D	b	1172	A C A T G G G G C C A C A A G G G G A C O C T A C T G A T T T C C T T A T A C C G C T G H C G A A A	1231
O	y	1274	t c c a a g a t a t g t t g g c c t g t g c t g t c t g t c t g c a a t a t t t g c c c g g c g c t a t t c a t g	1333
D	b	1232	T C C C G G G G C C T T A T A G C C C T A T C A C A T T A C C G C G T G G C G C A T T A C C C A T G G	1291
O	y	1334	c a a c t g c c c t c t o c t g g y a g a g t t c c t t c t t a t g a c g t a c g t g a c c c a a g a c t	1393
D	b	1292	C C G T C T C A A A T T T G T T G G C G G G G C A C C T G C C T G C T A T T - - - T T A T C T A C C T G A C	1348
O	y	1394	t g t a t t a t t g t a c a g t c c t g t g a t g t g g g a a g t t t g a g t a c a g c g t g c t t t	1453
D	b	1349	T G C A C T G G T T A A A C A T A T A T A C A T G T G T T G G C G A A T G G A A T G A C A T T G C A A T A C	1408
O	y	1454	c o a t g t c t a g t g t a a c a a g c o g a g t a t c c c a c a g t g t g t g a a a c a t g g t t g r	1513
D	b	1409	A A A T A T T G C C T G T G A T G C T A G T A T A C C C A C A T T G C H A G A A C C T C A A G T G A	1468
O	y	1514	g a g t a a g t c c a a g a c a t c c c g c c t g g g a a c a t c t g t c t c c a c t t c g t t t a c - - - c	1570

Dd	1469	TGtGTGTCTCCCTCCCTGtGTACATAGTGGGAATATACACCCTTCATATAGTCTCAGGC	1528	
Oy	1571	tgtgtcctacgaacgccttcctgcccatacttcatactatggaagtcgtacacccctcgacg	1630	
Dd	1529	TGAGGAGAGTCTGGCAACACTTGCCCCCATTTGTTGTGGCGTGTGGCGCTGCTTGGCCG	1588	
Oy	1631	ccatccctacccttgtttctcccaagaagcttcggctaccoccatccacagacacat	1685	
Dd	1589	CGGAGTACACCTACTTCTTCCAGAGACCAAGGGGGAGCGTTTGCACAGACCACT	1643	
RESULT 2				
US-09-040-444-5				
Sequence 5, Application us/09040444				
Patent No. 6063766				
GENERAL INFORMATION:				
APPLICANT: Koepsell, Hermann				
APPLICANT: Grundeman, Dirk				
APPLICANT: Gorboulev, Valentin				
TITLE OF INVENTION: Transport protein Which Effects The				
TITLE OF INVENTION: Transport Of Cationic Xenobiotics And/or Pharmaceuticals				
TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.				
NUMBER OF SEQUENCES: 6				
CORRESPONDENCE ADDRESSES:				
ADDRESSEE: Finegan,Henderson,Farabow,Garrett & Dunner, L.L.P.				
STREET: 1300 I Street, N.W., Suite 700				
CITY: Washington				
STATE: D.C.				
COUNTRY: USA				
ZIP: 20005-3315				
COMPUTER READABLE FORM:				
MEDIUM TYPE: Floppy disk				
COMPUTER: IBM PC compatible				
OPERATING SYSTEM: PC-DOS/MS-DOS				
SOFTWARE: Patent In Release #1.0, Version #1.30				
CURRENT APPLICATION DATA:				
APPLICATION NUMBER: US/09/040,444				
FILING DATE: March 18, 1998				
CLASSIFICATION:				
ATTORNEY/AGENT INFORMATION:				
NAME: O'Connor, Steven P				
REGISTRATION NUMBER: 41,225				
REFERENCE/DOCKET NUMBER: 2481,1453-01				
TELECOMMUNICATION INFORMATION:				
TELEPHONE: (202)408-4000				
TELEFAX: (202)408-4400				
INFORMATION FOR SEQ ID NO: 5:				
SEQUENCE CHARACTERISTICS:				
LENGTH: 1885 base pairs				
TYPE: nucleic acid				
STRANDEDNESS: single				
TOPOLOGY: linear				
MOLECULE TYPE: DNA (genomic)				
US-09-040-444-5				
Query Match	10.0%;	Score 183.8;	DB 3;	Length 1885;
Best Local Similarity	49.6%;	Pred. No. 2.8e-36;		
Matches 622; Conservative	0;	Mismatches 612;	Indels 21;	Gaps 5;
Oy	434	accttgaggagactgtgacgagaaagtctctcgaatgactgtggaattctgtaaggagtct	493	
Dd	407	ACAGAGACCACCTGCCCGCTGGGTCCCTCCACAGATGCGCTGGGTG-----ATGACACGC	460	
Oy	494	accttgccaccatlgtagccgagtgaaacctggtgtgtgtaggaagcactggaagccccac	553	
Dd	461	CCGGCTCTCCATCGTCACTGACTGATTCAACCTGGTGTGTGCTGACACCCGGAACCTGGACC	520	
Oy	554	tacaatatcctctgttcttcgttggtgtggtgcgttggtggctctctaattccaagagagcgt	613	
Dd	521	TCTTTTCAGTCCGTTTGAATGCGGGCTTCTCTTTGGCTCTCGGTCTTGACTATTG	580	
Oy	614	cagaacagtttgccggaagaatgtgctgttcgttgaccatbggatgagacagagcttca	673	

Db 581 CAGACAGGTTGGCCGTACAGTGTCTCCTGGGAACTGTGTCAGTCAACGCGGTGCGG 640  
QY 674 gctcccgagatctcttgagaatttgagatgttgcgycgttcttctgtg 733  
Db 641 GCGTGCATATGCGCTTCTGCCCCAACTACATGTCCATGCTGCTCTTCCGCTCTGAGG 700  
QY 734 gcatggcagatctccaaatgatgtgagacattgtcctgggagacaaattcttgac 793  
Db 701 GCCGTGCACGAAAGGCACTGATGCTGCTGCTACACACCTAATCAGAGATTGTGGC- 759  
QY 794 agtcagtcgataatactctctacgtttagagatgtgacataattatgcatcttgac 853  
Db 760 --TCGGGCTTCAGAAAGAGCGTGGCATCATGACAGATGCGCTTCACGCTGGGCTGG 817  
QY 854 tgggtgcgcacagcttctgacttacttaccgagacgtggagatgctgtgtgctgtga 913  
Db 818 TGGCGCTTACCGGGCTGCGCTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 877  
QY 914 cgaatgcgggggt 973  
Db 878 CCGTGGCCACCTTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 937  
QY 974 tcaatctcagggagcagattggaagagcagagatgacatcgcgaagctcgcgaagcca 1033  
Db 938 TGTATATCAAAAAGAAACACATGACATATAAAGATATGAGCACACATGCTCAAAAGA 997  
QY 1034 atggagatgt 1093  
Db 998 ATGGGAAGTGGCTCTGCTGATTTAA-----AGATGCTTTCCTCGAAGAGATGTCA 1051  
QY 1094 agaagcagcagctcccaacaattctgtatctgtctcgaacctggaatatccgagatgtca 1153  
Db 1052 CCGAAGAGCTGAGGCTTCAATTTGACACACCTGTTCCGACGCGCGCTGAGGAGCGCA 1111  
QY 1154 ccatcatgtcacaataatgt 1213  
Db 1112 CCTTCATCTGATGTACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1171  
QY 1214 atactctactgtcagtgagacatcttltgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1273  
Db 1172 ACATGGGCGCCACGAGCGGAACCTCACTGATTTCTTAACTCCGCTGCTGCTGCTGCTG 1231  
QY 1274 tcccaagataatgt 1333  
Db 1232 TCCCGGGGGGCTTCAATAGCCCTCATACCATGTAGCCGCGGCGCATATACCCCATGG 1291  
QY 1334 ccaatgcctctctcctgt 1393  
Db 1292 CCGTGTAAATTTGTGTGGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1348  
QY 1394 tgttatttggctacagctcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1453  
Db 1349 TGCATCTGTAAACATCATATATCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1408  
QY 1454 ccatgtctacgtgttacacagcagagctgtatcccaagctgtgtgtgtgtgtgtgtgtgt 1513  
Db 1409 AAATGATCTGCTGT 1468  
QY 1514 gagtcaagctcacaagcatcccgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1570  
Db 1469 TGGTGTGTCTCTCCCTGT 1528  
QY 1571 ttgtgtctagacagcttctgt 1630  
Db 1529 TGAGGAGAGTGTGGCAAGCTTGGCTCATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1588  
QY 1631 ccatcctacagcttcttctcccaagagcttgcgttacccacatcccaagacacat 1685  
Db 1589 CGGGAGTGACGCTACTTCTTCCAGAGACCAAGGGGAGCGCTTGTCCAGAGACCAT 1643

RESULT 3

US-08-592-126-109  
; Sequence 109, Application US/08592126  
; Patent No. 5821091  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Doljanov  
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
; NUMBER OF SEQUENCES: 151  
; CORRESPONDENCE ADDRESS:  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592.126  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 481 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: H993.seq  
US-08-592-126-109  
Query Match 8.7%; Score 158.4; DB 1; Length 481;  
Best Local Similarity 96.4%; Pred. No. 3.1e-30;  
Matches 162; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 618 caggttggcgggaagaatgt 677  
Db 169 CAGGTTGGCCGGGAAGATGTGCTGTGCGCATGGGACATGACAGAGGCTTACGCTT 228  
QY 678 cctgcagatctctcgaagaatttgaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 737  
Db 229 CCGCAGATCTCTCGAAGAATTTGAGATGTTTGTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 288  
QY 738 gggccagatctccaactatgt 785  
Db 289 GGGCAGATCTCCAACTATGTGGCAGCATTTGTCTGGGTATGGCAT 336  
RESULT 4  
US-08-592-126-108/c  
; Sequence 108, Application US/08592126  
; Patent No. 5821091  
; GENERAL INFORMATION:  
; APPLICANT: Gregory Doljanov  
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory  
; NUMBER OF SEQUENCES: 151  
; CORRESPONDENCE ADDRESS:  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592.126  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 4600-0111  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 481 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: H993.seq  
US-08-592-126-109

```
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,126
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Shultz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ. ID NO: 108:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: G78.seq
US-08-592-126-108

Query Match      8.5%; Score 155.2; DB 1; Length 370;
Best Local Similarity 95.2%; Pred. No. 1,8e-29;
Matches 160; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 618 caggttggccggaagaatgtcgttcgtgacatggcgatcagaagcttcagctt 677
    |||||
DB 221 CAGGTTGGCCGGAAGATGTGCTGTCGTGACCATGGCGTGCAGCAGGCTTCAGCTT 162

QY 678 cctgcagaatcttcgagaagaatttgagaatgttgcgtgtcttgccttgaagcat 737
    |||||
DB 161 CCTGCAGAACCTTCGGAAGATTTTGAGATGTTTGTGCTGCTGTTGCTCTTGTAGGCAT 102

QY 738 ggagcaaatctccaactatgttgagcattgttcctctgggagcaagaat 785
    |||||
DB 101 GGGCCAGATCTCCAACTATGGGCGACGATTTGTCTGGGTATGGCCAT 54

RESULT 5
US-08-501-572-4
; Sequence 4, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which Effects The
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flanagan, Henderson, Farbow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/501,572
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Toohay, Kimberlin M.
REGISTRATION NUMBER: 35,391
REFERENCE/DOCKET NUMBER: 02481.1453-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)408-4400
TELEFAX: (202)408-4400
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1882 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-501-572-4
```

```
Query Match      8.5%; Score 155; DB 3; Length 1882;
Best Local Similarity 47.4%; Pred. No. 3,8e-29;
Matches 566; Conservative 0; Mismatches 615; Indels 12; Gaps 3;
```

```
QY 499 tccaccattgacccgagtggaacctggtgtgtgaagcagatgaagcccaactaca 558
    |||
DB 434 TCCTCATGTCACGTAGATTGATTAACCTGGTGTGAGAGCCTGGAAGTGAGACTTTT 493

QY 559 atccctgttcttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 618
    |||||
DB 494 CAGTCTGTGTAACTGGGCTTCTTCCGTGGGCTCCTGTTGTTGATTTACATTCAGAC 553

QY 619 aggttggccggaagaatgtcgttcgttgacatggcgatgcagacgggttcagcttc 678
    |||||
DB 554 AGGTTTGGCGGTAAGCTGTCTGTTGGTGAACGAGTGATCATCTGTGTCGGTGTG 613

QY 679 ctgcagaatcttcgagaagaatttgagaatgttgcgtgtgtgtgtgtgtgtgtgt 738
    |||
DB 614 CTACACGCGGTGGCCCAACATATACATCCATGTTGCTCTTCGCTCTGCAGGGCATG 673

QY 739 ggccagatctccaactatgttgacgattgttcctggggagcaagaatcttcggaagca 798
    |||
DB 674 GTACAGCAAGGCAAGCTGGGTGTCGGCTATACCTTGATCACAAGATTTGTGGC---TCT 730

QY 799 gtctgataataatctctciacgtlaagagtgatcatatttatagtcatttgcattgtgataatgtg 858
    |||
DB 731 GGCTACAGGAGAAGACAGCGGCAATTTTGTACAGATGGGCTTCACAGTGCGGCTAGTGGGG 790

QY 859 ctgcacatgttgccttacttccagagatgggagatgcgtgcgtgtgtgtgtgtgtgt 918
    |||
DB 791 CTTGGCCGGGGTGCCATGCACTTCACAGACGTCGCTGCCTCCACATCTGCTGTGCTCGT 850

QY 919 ccgagggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 978
    |||
DB 851 CCTACCTTCCTCTTCTGTGTGTATACGTGTTGTGTCACAGATCCGCCGCTGTGTG 910

QY 979 tctcagggagcatttgaagagcagaggtgacatccgcagaggtgtccaaagccaatgg 1038
    |||
DB 911 TCCCAAGAGAGAACACGACGAGCTGTCAAGAAATAGAGCAAAATGTGCACAGAGAAGCGG 970

QY 1039 attgttgtcttccacatcttgaaccgggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1098
    |||
DB 971 AAGGTGCTCTCTGCTGACCTGAA-----GATGCTTGCCTTGAGAGAGATGCTCAGAA 1024

QY 1099 cagcagtcaccaacaacctctgtctcgaacctggaatatccgatgtgtaaccatc 1158
    |||
DB 1025 AAGCGAAGTCTGTGTTTGGCGACCTGTTCGCACTCCCAACCTGAGAGAACACACCTGC 1084

QY 1159 atgtcataatgtctgtgatgacataatcagttgggtatttgggtcttgcgtgataact 1218
```



Db	1085	ATCTGATGATCATCAATGAGTTCTTCTTGCTGCTGCTGTACCAAGGCTCTCATCATGACAGT	11444
QY	1219	cctaacttgcaatgaggagacactcttgtgaactgcttccttcacgcatglttgaagtcaca	1278
Db	1145	GGAGCCACAGGGGGCCAACTCTACACCTGAGACTTCTTTATTTCTTCTGATGGAAATTTCCC	1204
QY	1279	gcataatgltgtgacctgacgtgcctgcctgcgaatatattgccccggcgctatattcaatgacac	1338
Db	1205	GGCGCCTTCATCATCTCGTGACCAATTTGACCCGATTTGGCGGATCTATACCAATTTACCGGCC	1264
QY	1339	gacctcttcctggggtgcaagtgcctcttccttcacatcagctgtaacccccagactgcat	1398
Db	1285	TGCAATTTGGGAGAGGGGGGACGCTGCTGCTCTCATATCTTTATCCCGATGAGCTGCAC	1324
QY	1389	tatttgctacagtcctcgtgtgcatgagtgtgggcaattgtgaagtcacgctcttcctacatg	1458
Db	1325	TGGTTGAGACCTTACCCCTCGCTGCTCTTTGGCGGATGGGGGCCACCAATTTGCTGCAGATG	1384
QY	1459	gtcacaagtcaacagagccagcagctctcccaacagtgtagaacaatggtgtggaatc	1518
Db	1385	GTCCTGCTGGTGAACCTGTGAGCTGATGCCATCTTATATGGAATCTTGGATGATGAT	1444
QY	1519	agctccacagcatcccgccctcgggacagcatcctcctccacttcggttac---cttgt	1575
Db	1445	TGCTCTGCCCTGATGTAGACTGTGGGTGGGATCTTACCCCCCTTATGTGTTCAGCGCTATG	1504
QY	1576	gactaagacgctctcctgacctacattctcaatlyggaagctctgacatccctgacagcatc	1635
Db	1505	GAACTTTGGCAAGCCCTCGCCCTCATTTTGTTTGGGGTTTGGGCTGTGACTCTGGGGCC	1564
QY	1636	ctacactgtttctcccaagagacttcglttaacccactccagaacaacatgta	1688
Db	1565	ATGACCTCTCTCTCCAGAGACCAAGGGGTGGCTTTCCCTGAGCATTTTA	1617

RESULT 6  
 US-09-040-444-4  
 ; Sequence 4, Application US/0904044  
 ; Patent No. 6063766  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Koepsell, Hermann  
 ; APPLICANT: Grundeman, Dirk  
 ; APPLICANT: Gorboulev, Valentin  
 ; TITLE OF INVENTION: Transport protein which effects the  
 ; TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals  
 ; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.  
 ; STREET: 1300 I Street, N.W., Suite 700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/040,444  
 ; FILING DATE: March 18, 1998  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: O'Connor, Steven P  
 ; REGISTRATION NUMBER: 41,225  
 ; REFERENCE/DOCKET NUMBER: 2481.1453-01  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)408-4000  
 ; TELEFAX: (202)408-4400  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1882 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-09-040-444-4

```

Query Match	8.5%;	Score 155;	DB 3;	Length 1882;
Best Local Similarity	47.4%;	Pred. No. 3.8e-29;		
Matches 566;	Conservative	0;	Mismatches 615;	Indels 12;
				Gaps 3;

[illegible]

Db	6/4	GTCCAGCAAGGGCACACTGGGGTGGTCGGCGCTATACCTTTATCCAGAGATTTGTGGC---	-GCT	730
Oy	799	gttcgataaatattctctaaagtagagtgatgcatattttaaagcttggccaatgtagt		858
Db	721	GGCTACAGAGAAACAGACGGGCATTTTGTACCAAGATGGCCTTCACAGTGGGCGTAA		790
Oy	859	ctgcgaactgtttgcttaacttaatccagagactgcggaatgcgtctgtgtgcgtgaagat		918
Db	791	CTTCCCGGGGGGCGCTATGGCAATTTCCAGATGGCGCTGGCTCCACCTACCTGTGTCC		850
Oy	919	ccggagagtgctgtcgtgcagcatctgtgtgttcatccctgaatcccccagatgcatc		978
Db	851	CTTACTTCCTCTTCGTCGTGTATTACTGGTTTGTGCCAGAAATCCCGCGGTGGTGG		910
Oy	979	tctcaagggaagatttgaagaagcagaagtgatcatatccgaaggtgccaaagccaaatgg		1038
Db	911	TCCACAGAGAAACACACGCGAGCTGTGAGATATATGACCAATTTGCACAGAAAGACGG		970
Oy	1039	atgttgctgccttcaactatcttgaacccagtagagttcaagaaactaaagtccaagaag		1098
Db	971	AAGGCGCTCTCTGCTGACGTGA-----GATGCTCTGCTTGAGAGAGATGCTTCAGAA		1024
Oy	1099	cagagatcccaacaacatctctgatactgttgacttggaacttgaaatccgaatgtagccact		1158
Db	1025	AAGGAAGTCTCTTGTTGGCGACCTGTTCGACATCCCAACCTGAGAGACACACCTTC		1084
Oy	1159	atgtccataabtcgttgaatgagacataatcagtgagctaatlttggagcttcgctgataat		1218
Db	1085	ATTCGATGATGTATATATAGTTCCTGTGCGTGTCTGTACCAAGGTCCTATCATCGACGG		1144
Oy	1219	cctaacttgatgagacatcttltggaactgttcccttaagcagatgttgaagtcca		1278
Db	1145	GGAGCCACAGGGGCCAACCTCTACCTGGAATTTTATTTATTTCTCTGGTGAATTTCCC		1204
Oy	1279	gcataatgttbgaccltgcgtcgtcgttgcataatltgcccgogcatatccatctgagccact		1338
Db	1205	GGGGCTTCATCATTCCTGTGCACCAATGACCGCATTTGGCCGATCATCCCAATACCGGCC		1264
Oy	1339	ggccctttccctgggtgcagtgctcttccttcctatbcagctgtgatcccccaagacttgat		1368
Db	1265	TGCAATCTGGTGAACGGGGGCAACCTGCTCTCTATATATCTTATATCCCGATAGCTGGAC		1324
Oy	1399	tatttgctacagtcctcgttgatgtgtgtggcagaatttggagtaacagctgtcccttccatg		1458







```

ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

```

```

Query Match
Best Local Similarity 6.2%; Pred. No. 0.0012;
Matches 26; Conservative 218; Mismatches 173; Indels 0; Gaps 0;

```

```

QY 1285 ggtgtgagcctgctgctgctgcaaatlgtccgcgcgtatccagtcacgtgcctc 1344
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1056 GAGCTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1115

QY 1345 ttctctgggtgagcagtgctctctctcaatgacgtgtaacccagacattgatttg 1404
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1116 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1175

QY 1405 gctacagtcctggtgagtgagtgcaagtgagtgacgctgcttcttcacagtgctac 1464
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1176 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1235

QY 1465 gtgtacacagcagcagctgtatccacagtggtgagaaacatggtggtgagtcac 1524
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1236 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1295

QY 1525 acagcaccgcctgcgcgcagacatcgtctccctactcgtttacacttgctgctacgac 1584
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1286 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1355

QY 1585 cgtctccgcctacatctcatctgaggaagctgacacatccctgacagcattcacttg 1644
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1356 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1415

QY 1645 ttctccagagagcttcgtaccccaactccacacacattgacagatgtaaga 1701
    : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1416 YYYYYYYYYYYYYYYYYYGTACCAAAATCTTATCTCTTTTACACTGTCANAGA 1472

```

RESULT 11

```

PCT-US95-14418-3
; Sequence 3, Application PC/TUS9514418
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: DNA Encoding a Thermostable DNA Polymerase Enzyme
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14418
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gass, David A.
; REGISTRATION NUMBER: 38,153
; REFERENCE/DOCKET NUMBER: 28003/32330
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1794 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1794
; PCT-US95-14418-3

```

```

Query Match
Best Local Similarity 46.5%; Pred. No. 0.0011;
Matches 205; Conservative 0; Mismatches 233; Indels 3; Gaps 1;

```

```

QY 94 ggcgcgctctgtggtgctctgagggcgagcatgagcagtaacagagtgacgccttc 153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 337 GTCGGGGGCTCTCTCGCAAGGACCTCGCCGCTTGCGCTCGAGGAGGCGCTAGACCTC 396

QY 154 ctggcgagatgggggacctccacagcgtcatcttctctgctcagcgacgacatc 213
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 397 GTGCCCGGAGACGACCCCAAT--GCTCTCGCTTACTCTCTGAGCCCTCCCAACACAC 453

QY 214 cccaatgagctcagcgagcgtctgctcgtgtctctgatatgagaccggagagcgcctgc 273
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 CCGGAGGGGTGGCGCGCGCTTACGGGGGAGTGAACGAGAGACGCCGCCACCGGGGCC 513

QY 274 cgggtgacgagcgcgcgagacatgagcagcgtctgagcaaacacactgttccactgcyg 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 CTCCTCTCGAAGAGGCTCATCGAAGCTCTTAAGCGCGCTCGAGGGGAGAGAGAGCTC 573

QY 334 ctggcgagagcgccgcgagatgtcccaacagctgcgcgctacgcgtgcacatcgcc 393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 CTTTGCTCTACACGAGGTGAGAAAGCCCTCTCCGGGCTCTGAGCCACATGAGAGGC 633

QY 394 aactctcgagcgtcggtgagccggggcgagacgtgagcttgaggagagagagag 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 ACCGGGATACGCTGAGAGTGGCTTACCTGACAGGCCCTTTCCCTGAGACTTGGAGAG 693

QY 454 gagagctgtcgtgagtgctggaggtcagtcacagcgttaccactgtgaccattgtgacc 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```





Qy 328 ctgcggtctgggacgagcggtgccccacagctgcccgcgtaccggctcgccacc 387  
|| | | | | | | | | | | | | | | | | | | | |  
Db 1860565 ctgcgcaaccgctcatcgaccacgtgcggaacgagactgctgttggcatcgcgctg 1860624  
|| | | | | | | | | | | | | | | | | | | | |  
Qy 388 atcgccaattctcggtcggtcggtgagccggggcgcgacg 430  
| | | | | | | | | | | | | | | | | | | | |  
Db 1860625 accgccccgcagcgaagagagccagggcaccgcccgcgacg 1860667  
| | | | | | | | | | | | | | | | | | | | |

Search completed: July 17, 2002, 00:27:42  
Job time: 15613 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 16, 2002, 21:46:42 ; Search time 4179.87 Seconds

(without alignments)  
5912.367 Million cell updates/sec

Title: US-09-521-195B-4

Perfect score: 1831

Sequence: 1 cggagcgcgtcttggtgcgcct.....agggaagaactgaagagaa 1831

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues 27472414

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pin:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	657	35.9	754	10	BG332869 602430714
2	653.2	35.7	910	10	B1552541 603194456
3	617	33.7	905	9	AL557809
4	601.4	32.8	611	10	BG723644
5	585.4	32.0	689	10	B1547770
6	573.4	31.3	807	10	BE546219
7	559.6	30.6	799	10	B1647860
8	547	29.9	806	10	B1461374
9	486	26.5	800	10	B1683731
10	466.8	25.5	846	10	BE783848
11	463.8	25.3	917	10	BE783863
12	460.2	25.1	694	10	B1758502
13	455	24.8	543	10	BE078340
14	449.6	24.6	584	10	BG087364
15	449.2	24.5	667	10	B1143364
16	445.4	24.3	641	9	BB615168
17	445.2	24.3	640	9	BB615170

18	427.2	23.3	879	10	BG870249
19	422.2	23.1	522	10	BE230186
20	420	22.9	645	9	BB617187
21	418.6	22.9	762	10	BE782868
22	417.2	22.8	665	10	BB614876
23	412	22.5	806	9	BG961802
24	409.4	22.4	988	10	BE732425
25	405.2	22.1	922	10	B1251565
26	401.2	21.9	482	10	BM032317
27	394	21.5	484	10	BE774599
28	390	21.3	935	10	BG242987
29	376.4	20.6	580	10	B1339311
30	346.8	18.9	1004	10	BG469138
31	330.2	18.0	839	10	BE784044
32	327.4	17.9	670	10	BJ041781
33	325.6	17.8	877	10	B1110037
34	324.4	17.7	735	10	BJ068088
35	322.6	17.6	353	10	B1763685
36	321.4	17.6	430	9	AW909784
37	319.2	17.4	606	10	BE152158
38	308.6	16.9	395	9	AM484092
39	302.4	16.5	928	9	AL557808
40	301.4	16.5	550	9	AL645203
41	289.6	15.8	726	10	BG920021
42	271	14.8	578	10	BE846144
43	269.6	14.7	421	9	BB846601
44	268.2	14.6	453	10	BE567603
45	267	14.6	612	10	BE862866

## ALIGNMENTS

RESULT 1  
BG332869 754 bp mRNA linear EST 27-FEB-2001  
LOCUS 602430714P1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4548603 5',  
DEFINITION BG332869  
ACCESSION BG332869  
VERSION BG332869.1 GI:13139307  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 754)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DPF/Gazdar  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.lnl.gov>  
Plate: LINC1238 row: k column: 04  
High quality sequence stop: 741.  
Location/Qualifiers  
1. 754  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4548603"  
/clone\_id="NIH\_MGC\_18"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pORF7; Site: 1; XhoI: Site: 2;  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(g). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,

## FEATURES

source





```
/db_xref="taxon:9606"
/clone="IMAGE:4829882"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/Note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGR), National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT      96 a      194 c      201 g      120 t
ORIGIN

Query Match      32.8%; Score 601.4; DB 10; Length 611;
Best Local Similarity 99.0%; Pred. No. 8.5e-116;
Matches 605; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 88 acccagcgccgctctgtgagcctctgagggcgcaatgcgagactacagcaggtgacc 147
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 AGCGAGGCGCCGCTGTGTGGCTGTGAGGGCGCATGCGGACTAGAGAGTGAC 60

QY 148 gectctgtgagagtgaggagccctccagcgcctcatctctctcgtcagcgagc 207
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 GCTTCCCTGGGCGAGTGGGGGCGCTTCCAGCGCTCATCTTCTTCCGTGAGCGCCAGC 120

QY 208 atcatcccaatgcttaccagcgcctcctcgtcttctcgtatagaccccgagacac 267
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 ATCATCCCATATGCTTACCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180

QY 268 cgtctgcgggtgcgcgagccgcgacactgagcagcgcttgcgcaacacactgtccca 327
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 CGTGTCCCGGTGCGGAGCGCCGACACTGAGCAGCGCTGCGCAACACACTGTGCCA 240

QY 328 ctgagcgctgcggagcgagcgaggtgtgcccacagccttgcgcccgtcagctgcgac 387
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 CTGCGGCTGGCGGAGCGCGGAGGTGCGCCACAGCTGCCCGCTGACGCGCTGCCACC 300

QY 388 atcgcaactctcgcgcgtcgcgagctgagacccgagcgcgagcgtgcgacttgcgagctg 447
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 ATCGCAACTTCTGCGGCTGTGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAGCGCG 360

QY 448 gagcagagagctgtctgtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 507
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 GAGCAGGAGAGCTGTGTGATGCTGAGGAGGTTCAGTCAAGTCAAGTCAAGTCAAGTCA 420

QY 508 gtgacgagtgagaaactgtgtgtgagagacacttgaagagcccaactacacatctcttg 567
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 421 GTGAGGAGAGTGAACCTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480

QY 568 ttctctgtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 627
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 481 TTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540

QY 628 cggagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 687
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 541 CGGAAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600

QY 688 ttctcgaagaa 698
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 601 TTCTCGAAGAA 611

RESULT 5
BIS47770 689 bp mRNA linear EST 05-SEP-2001
LOCUS BIS47770 603191940P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5263353 5',
DEFINITION mRNA sequence.
ACCESSION BIS47770
VERSION BIS47770.1 GI:15435082
```

```
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cggabbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LLM11663 row: p column: 10
High quality sequence stop: 689.

FEATURES
source
1..689
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5263353"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/Note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGR), National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT      92 a      244 c      231 g      122 t
ORIGIN

Query Match      32.0%; Score 585.4; DB 10; Length 689;
Best Local Similarity 99.7%; Pred. No. 2e-112;
Matches 597; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 cggagagcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 91 CGGACGGTCTTGGGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 150

QY 61 cagcgcaaaagccgcgcgcttccagacccagagcgcgctctgtgagcctctgagggc 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 151 CACGGCGAAAGCCCGCGGCTTCCCGACCGCGCGGCTGTG38CTGTGAGGGC 210

QY 121 ggcattcgaggaactaagagagtgagccgcttctgtggagtgagtgagtgagtgagtgag 180
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 211 GGCATGCGGGAGCTACAGACGAGTGACCGCTTCTGCGGAGTGGGAGTGGGAGTGGGAG 270

QY 181 ctcatcttctctgtctcagcgcagacatcatcccaatgtgtgtgtgtgtgtgtgtgtgtgt 240
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 271 CTGATCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 330

QY 241 gttctctgaatagaccccgagacacgctgcgggtgtgcggaacgcgcggaactgagc 300
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 331 GTGTCTCTGATAGGAGACCCCGGAGACCGGCTGCGGAGTGGCGGAGCCCGAAGCTGAGC 390

QY 301 agcgctgtgcgaacaacacatgtcccaactgtgcgagtgagcgagtgagtgagtgagtgag 360
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 391 AGCGCTGTGGGCAACACACTGTCTCCACTGCGGCTGCGGAGCGCGGAGGTGCGCCAC 450

QY 361 agctgcgcgcctacacggtctgcacacatgcgcaacttctcggcgctcgggtgtgagcg 420
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 451 AGCTGCGCGGCTACCGGCTGCGCACCATGCGCAACTTCTGCGGCGCTGCGGAGTGGAGCGG 510
```

QY 421 gggcgcagcgtgagccttgaggcagctgagcagagagctgctgagatgctggagctc 480  
 |||||  
 Db 511 GGGCCGACGtGAGACCTGGGGACCTGAGAGAGAGACGCTCTGAGATGGCTGGGATTC 570  
 QY 481 agtcagagcgtctaccctgctcacacattgtgacgagtggaacctggtgtgagga-cga 539  
 |||||  
 Db 571 AGTAGGAGCGTCTACCTGTCTCCACCATGTGTACCGAGTGGAACCTGTGTGAGAGACGA 630  
 QY 540 ctggaagggccacctcaaatctctctgtctgtctgtgtgtgtgtgtgtgtgtgtc 598  
 |||||  
 Db 631 CTGGAAGGCCCACTCAACATCTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTCA 689

RESULT 6  
 BE546219 807 bp mRNA linear EST 09-AUG-2000  
 LOCUS 601071250F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:3457116 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE546219  
 VERSION BE546219.1 GI:9774864  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 807)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LHAM846 row: d column: 13  
 High quality sequence stop: 540.  
 Location/Qualifiers  
 1..807

FEATURES  
 source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3457116"  
 /clone\_lib="NIH\_MGC\_12"  
 /tissue\_type="cervical carcinoma cell line"  
 /lab\_host="DH10B"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.4 kb. Library prepared by Life  
 Technologies."  
 BASE COUNT 210 a 216 c 194 g 187 t  
 ORIGIN

Query Match 31.3%; Score 573.4; DB 10; Length 807;  
 Best Local Similarity 95.5%; Pred. No. 6.9e-110;  
 Matches 633; Conservative 0; Mismatches 26; Indels 4; Gaps 4;  
 QY 747 ctccaactatgtagcagcattgtctctggggaagaatctctggcaagtcagttcgtat 806  
 |||||  
 Db 1 CTCCAACTATGTGGACACATTCCTCGGGGAGACGAATTCCTGGCAAGCTTCGTAT 60  
 QY 807 aatatctacgttagagatgcatatttatgcatgtgctacatggtgtgcac- 865  
 |||||  
 Db 61 AATATCTCTACGTTAGGAGTGTGCATATTTTATGCAATTGGCTACATGCTGCTCCACA 120  
 QY 866 tgttcttactcaccgcagagctggcgagatgctgtgtgtgtgtgtgtgtgtgtgtgt 925  
 |||||  
 Db 121 TGTTCCTACTTCATCCGAGACTGGCGGATGCTGCTGTGTGTGTGTGTGTGTGTGTGT 180

QY 926 tgcgtgcgtggacactgtggtttacatccctgagatcccccgatgagtcacatctcagg 985  
 |||||  
 Db 181 TGTATATGCGTGGACACTCTGGTGTATCTCCCTGAGTCCCGAGTGGCTCATCTCTCAG 240  
 QY 986 gacgattgaagagcagagtgatcatccgcgaagctcccaagccaaatggatgtgtg 1045  
 |||||  
 Db 241 GACGATTTGAAGAGGAGAGAGGTGATCATCCGCAAGGCTCCAAAGCAATGGGATCGTAG 300  
 QY 1046 tgccttccactatcttgaccagagtgatgaacaaactaagttccaaagaagcagat 1105  
 |||||  
 Db 301 TGCCTTCCACTATCTTTGACCCGAGAGTATCAAGACCTAAGTTCAGAGAGCAGAGT 360  
 QY 1106 ccacacacttcgtgactgcttcagacctggaataatccgagttgtacaccatgltca 1165  
 |||||  
 Db 361 CCCACAACATCTGGATCTGCTTCGACCACTGGAAATCCGGATGGTCCACCATCATGTCCA 420  
 QY 1166 taatgtctgtagtaccatacaagtgagctatcttgagcttcgtctgatactcctaact 1225  
 |||||  
 Db 421 TAATGCTGTGATGACCATATCATGAGTGGCTATCTGGGCTTACGGTGTATCTCTTA 480  
 QY 1226 tgcgtgagacatcttctggaactgtctccttcagcgatgtgtgaagtcacagatgt 1285  
 |||||  
 Db 481 TGCATGGGACATCTTGATGTAACCTGCTTCAACCGATGGTCCAGATCCAGCATATG 540  
 QY 1286 tgt-tgcctgtgtgc-tgtcgaatatctgcccgcgcgtatctcatggcagctgcct 1343  
 |||||  
 Db 541 TGTAAAGACTGGCTGCTTCTGCAATATAGCCCCGGCGCTATTCCATGCGCATGACCT 600  
 QY 1344 ctctctgtgtgcaagtgctctctctcttcaatgca-gctgtgtacccccagactgtat 1402  
 |||||  
 Db 601 CTTCCTGGGTGGACAGTCTCTCTCCATGCACGCTAGTACACAGCATGTCTGACT 660  
 QY 1403 tgg 1405  
 |||||  
 Db 661 AGG 663

RESULT 7  
 B1647860 799 bp mRNA linear EST 12-SEP-2001  
 LOCUS 603276129F1 NIH\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5316559 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1647860  
 VERSION B1647860.1 GI:15562096  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 799)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: lochar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LHAM11802 row: i column: 08  
 High quality sequence stop: 736.  
 Location/Qualifiers  
 1..799

FEATURES  
 source  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5316559"  
 /clone\_lib="NIH\_CGAP\_Mam3"  
 /tissue\_type="tumor, gross tissue"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: Salt. Cloned unidirectionally. Primer: Oligo dT.  
Average Insert 2 kb. Library constructed by Life Technologies, catalog #12017-018. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999). Note: this is a NCI-CCGAP Library."

BASE COUNT 172 a 198 c 215 g 213 t 1 others  
ORIGIN

Query Match 30.6%; Score 559.6; DB 10; Length 799;

Best Local Similarity 85.3%; Pred. No. 5, 4e-107;  
Matches 689; Conservative 0; Mismatches 110; Indels 8; Gaps 5;

QY 346 cgcgaagtgccccacagctccgcgcctaccgcgtgcgcacatcgcaactctcgcg 405  
D 1 CGACAGGCTGCTCAGAAATGCCCGCTACCGATGCGCCACCATCGCACTTCTCTGAG 60  
QY 406 ctggagcttgagcgccggcgccgagcttgagcttgagcgagcgagagctgctg 465  
D 61 CTAGGGCTGGAGCCGGGGCGGAGGTGGACCTGGAGCAGCTGGAGCGAGAGCTGCTG 120  
QY 466 gatggcctggagctcagtcagagagcttacctgtccaccactgtgaccgagtggaactg 525  
D 121 GATGGCTGGAGGTACGACAAAGGAGCTTCTTCCTGTCACCATCGACAGAGTGGAGACTG 180  
QY 526 gtgtgtgaggaagcagctggaagccccacatcaactccctgtctctcgtggtgtgctg 585  
D 181 GTGTGTAAGGATGACTGGAAAGCCCACTACCACTCTGTTTTCGGGGGTGCTG 240  
QY 586 ttggagctccctcatcttcaaggcagctcagacagagcttgccggaagatgtgctgttc 645  
D 241 ATGGGCTCTTCATTTTCAGGACAGCTCTCAGACAGG-TTGGTCCCAAGATGTGCTGTT 299  
QY 646 gtgacatgagcagtcagcagcagctcagcttctctcagacatctctcgaaatlttgag 705  
D 300 TTGACATGAGCGCATGACAGACTGCTTCAGCTTCTCGACAGGTCTCTGTGAACCTTGAG 359  
QY 706 atgtttgcgctgctgttctctgttgagcagtcgagagcttccaaactctgagcagca 765  
D 360 ATGTTACAGTGTCTTTTGTCTTGTCGTCAGTGGGTGATCTCAACTACGTGAGCA 419  
QY 766 ttcttccttgaggaagaatctctgcaagtcagctcgtataataatctctacgttaga 825  
D 420 TTTGCTCGGAGACAGAAATCTTTCCAGTCATTCGAATTAATTCGCCACTTAGA 479  
QY 826 gtgtgacataatttatgcatcttgctacaatggtctgcacagt-ttgcttaactcatcg 884  
D 480 GT-TGCATATTTTATGCTTTGGCTTTCATGCTGCTGCACAGTCTTCATACCTCATCAG 538  
QY 885 agactggcgagatgctgctgctgctgctgcagcagtcgcgggggtgtgtgctgagcactg 944  
D 539 AGACTGGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 598  
QY 945 gtgtttcatccctagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 1004  
D 599 GTGGTTCATCCCTGATGCCACAGATGCTCATCTCAAGCGCGAATTAACAGAGCAGAGA 658  
QY 1005 ggtgatactcc--gcaagctgcgcaagcgaatggtgattgtgtgcttccacatctt 1062  
D 659 GGTGATCATCCCGCAAAAAGCTGCAAAAATCAATGGGATGTGTGACCTTCACATATCTTC 718  
QY 1063 gaccgaagtgagttacagagcttaagttccaa--ggaagcagagctccacacacatctg 1119  
D 719 GATCCAAATGAGTTTCAAGAAATTTAAATTTTAAGGAAAGCTCAGTTGCAACCAATTTAT 778  
QY 1120 gatctgctcgaacctggaat 1140  
D 779 GATCTGATCGAACACGGAAT 799

RESULT 8  
BI461374

LOCUS BI461374 806 bp mRNA linear EST 21-AUG-2001  
DEFINITION 6032068381 NIH\_MGC\_97 Homo sapiens CDNA clone IMAGE:5272472 5',  
ACCESSION MRNA Sequence.  
VERSION BI461374  
KEYWORDS BI461374.1 GI:15252030  
SOURCE EST.  
ORGANISM human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 806)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Miklos Palcovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraiki  
Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LLM11687 row: 1 column: 09  
High quality sequence stop: 751.  
FEATURES  
source  
location/Qualifiers  
1..806  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5272472"  
/clone\_1b="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TGTGTTTTTTTTTTVN-3',  
size-selected for average insert size 2.2 kb and  
normalized to ROI 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 149 a 227 c 256 g 174 t  
ORIGIN

Query Match 29.9%; Score 547; DB 10; Length 806;  
Best Local Similarity 87.2%; Pred. No. 2, 4e-104;  
Matches 700; Conservative 0; Mismatches 25; Indels 78; Gaps 6;

QY 115 gaagcgagcagtcgagactacagcagagtgacgccttctctgagcgagtgaggcccttc 174  
D 6 GGGGGGGGATGGGGAGTACGACGAGTGACCGCTTCTGCGGAGTGGGGCCCTTC 65  
QY 175 cagcgctcaatcttctctctgctcagcgccaagatcatctcccaatggtttacccgctg 234  
D 66 CAGCGCTCATCTTCTCTGCTCAAGCGCAGCATATCCCAATGGCTTCACCGGCTG 125  
QY 235 tctctcggttctctgtagtagcagcccgagcgacccgctgctcgagcgagcgagcgaac 294  
D 126 TCTTCGTTGTTCTGTATGAGGACCCGGAGACCGCTGCGGGTCCGAGCCCGGAAC 185  
QY 295 ctgagcagcgctcgagcgaacacacactgtccactgtcgagctcgaggaaggcgagtg 354  
D 186 CTGAGCAGGCGCCGTGGGCAACACATGTGCCACTCGGCTGGGGAGCGCGCGAGGTG 245  
QY 355 ccccaagctgcgcgcgctacagcgctgcgcacatctgccaacttctcgcgctcgagctg 414  
D 246 CCCACAGCTGCCCGCTACCGCTGCGCACCATGCGCAACTTCGCGCGCTGGGCTG 305  
QY 415 gaagcggagcgagcagcttgacacctgggagcagcttgagcagagagatgtctgagtgctg 474  
D 306 GAGCCGGGGCGCAGCTGAGCTGGGGCACTGGAGCAGAGAGCTGTGAGATGGCTGG 365

[illegible]

[illegible]

QY	468	tgctgggagatctcaatcaggaagcttactatgcccacattgagaccattgaaacttg	527
Db	448	tgcctggggatgacgacaaagagctcttctgtgccaccatcggaagagatgggacctgat	507
QY	528	gtgtagagagcactggaagcccccactacaaatctctctgttcttcgtgaggtgtcgtc	587
Db	508	gtgttagaggtgctctggaaaagcccccactacaccctctctgttttttcgtgggggtctcat	567
QY	588	ggagcctctcatctttagagagcagctctcagacaggtttgcccgaagaatgtgcgttgct	647
Db	568	gggctt-ctttcattttagagacacactcttcagacacagttttgctgcgaagatgttctcttt--tt	625
QY	648	gaacatgagcatatgacagacagcttcactcttcctgcagatcctctcgaagaatttgat	707
Db	626	gaccatgggcatacgacagactgg--tttagcctcttcgcacactcttctctgtaacatttcgaat	684
QY	708	gtttgtctgctgttttctcattgtatgacatgggcagagctcccaactttgttgacat	767
Db	685	gttttagcag---gcctttgtccttctt--gcattggcttcattctccacttttagctggagcatt	739
QY	768	tgctcctg	775
Db	740	gggtccggg	747

[illegible]

FEATURES	Source	Location/Qualifiers
		1. 917
		/organism="Mus musculus"
		/strain="FVB/N"
		/db_xref="taxon:10090"
		/clone="IMAGE:4240241"
		/clone_lib="NCI CGAP Kid14"
		/lab_host="DH10B (TL phage-resistant)"
		/note="Organ: Kidney; Vector: pCMV-SPO06; Site.1: NotI; Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library. 1"
BASE COUNT	169 a	268 c 306 g 173 t 1 others
ORIGIN		
Query Match	25.3%	Score 463.8; DB 10; Length 917;
Best Local Similarity	82.0%;	Ref. No. 6.8e-87;
Matches 594; Conservative	0;	Mismatches 123; Indels 7; Gaps 5;









Wed Jul 17 08:31:26 2002

us-09-521-195b-4.rst

Db 659 TATCTCA 666

Search completed: July 16, 2002, 21:46:52  
Job time: 7988 sec